The Telecommunications and Data Acquisition Progress Report 42-106

April–June 1991

E. C. Posner
Editor

August 15, 1991

NASA
National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The Telecommunications and Data Acquisition Progress Report 42-106

April–June 1991

E. C. Posner
Editor

August 15, 1991

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Preface

This quarterly publication provides archival reports on developments in programs managed by JPL's Office of Telecommunications and Data Acquisition (TDA). In space communications, radio navigation, radio science, and ground-based radio and radar astronomy, it reports on activities of the Deep Space Network (DSN) in planning, in supporting research and technology, in implementation, and in operations. Also included is standards activity at JPL for space data and information systems and reimbursable DSN work performed for other space agencies through NASA. The preceding work is all performed for NASA's Office of Space Operations (OSO). The TDA Office also peforms work funded by two other NASA program offices through and with the cooperation of the Office of Space Operations. These are the Orbital Debris Radar Program (with the Office of Space Station) and 21st Century Communication Studies (with the Office of Aeronautics and Exploration Technology).

In the search for extraterrestrial intelligence (SETI), *The TDA Progress Report* reports on implementation and operations for searching the microwave spectrum. In solar system radar, it reports on the uses of the Goldstone Solar System Radar for scientific exploration of the planets, their rings and satellites, asteroids, and comets. In radio astronomy, the areas of support include spectroscopy, very long baseline interferometry, and astrometry. These three programs are performed for NASA's Office of Space Science and Applications (OSSA), with the Office of Space Operations funding DSN operational support.

Finally, tasks funded under the JPL Director's Discretionary Fund and the Caltech President's Fund that involve the TDA Office are included.

This and each succeeding issue of *The TDA Progress Report* will present material in some, but not necessarily all, of the following categories:

OSO Tasks:
- DSN Advanced Systems
  - Tracking and Ground-Based Navigation
  - Communications, Spacecraft–Ground
  - Station Control and System Technology
  - Network Data Processing and Productivity
- DSN Systems Implementation
  - Capabilities for Existing Projects
  - Capabilities for New Projects
  - New Initiatives
  - Network Upgrade and Sustaining
- DSN Operations
  - Network Operations and Operations Support
  - Mission Interface and Support
  - TDA Program Management and Analysis
- Ground Communications Implementation and Operations
- Data and Information Systems
- Flight–Ground Advanced Engineering
- Long-Range Program Planning

OSO Cooperative Tasks:
- Orbital Debris Radar Program
- 21st Century Communication Studies

OSSA Tasks:

Search for Extraterrestrial Intelligence
Goldstone Solar System Radar
Radio Astronomy

Discretionary Funded Tasks

Contents

OSO TASKS
DSN Advanced Systems
TRACKING AND GROUND-BASED NAVIGATION

COMMUNICATIONS, SPACECRAFT–GROUND

STATION CONTROL AND SYSTEM TECHNOLOGY

DSN Systems Implementation

CAPABILITIES FOR EXISTING PROJECTS

CAPABILITIES FOR NEW PROJECTS

DSN Operations

NETWORK OPERATIONS AND OPERATIONS SUPPORT

TDA PROGRAM MANAGEMENT AND ANALYSIS

Precise Orbit Determination of High-Earth Elliptical Orbiters Using Differenced Doppler and Range Measurements

J. A. Estefan
Navigation Systems Section

Recent advances in Deep Space Network station calibration methods have led to renewed interest in the use of differenced Doppler and range data types for interplanetary navigation. This article describes an orbit determination error analysis of the performance of these differenced data types when used with conventional two-way Doppler for precise navigation of High-Earth Orbiters. Three highly elliptical Earth orbits are investigated, with apogee heights on the order of 20,000 km, 70,000 km, and 156,000 km. Results indicate that the most significant navigational accuracy improvements, relative to the performance obtained from two-way Doppler alone, are achieved for the lowest altitude orbit by using differenced Doppler measurements with two-way Doppler (assuming that spacecraft onboard downlink antennas have no ground footprint limitation in the near-apogee regime). In the case of the two higher altitude orbits, accuracy improvements over Doppler-only performance, although less dramatic, are also achieved when differenced range measurements are combined with two-way Doppler.

I. Introduction

As NASA's commitment to supporting national and international High-Earth Orbiter (HEO) missions continues to evolve, it is becoming increasingly apparent that the NASA/JPL Deep Space Network (DSN) will be tasked with providing tracking and navigational support, particularly for the orbiting radio astronomy platforms planned for launch as early as 1995. The DSN is proceeding with the implementation of a dedicated HEO subnetwork of 10-m antennas for this purpose.[1]

A host of tracking techniques and data types have been investigated for possible use in HEO mission support. The most common data types that are expected to receive widespread use are two-way Doppler and two-way range. High-accuracy interferometric and differential techniques, such as very long-baseline interferometry (VLBI) and ΔVLBI, as well as Global Positioning System (GPS)-based tracking have also received much attention [1,2]. One technique that has been suggested is a differenced one-way range (DOR) by multiple frequency phase measurements, which modulates the transmission of the spacecraft so that cycle ambiguities can be eliminated and which measures only the phase differences [3] (not to be confused with ΔDOR, in which an observation of an ex-

[1] J. Ovnick, "Orbiting VLBI Subnet C/D Review," Presentation Viewgraphs, Jet Propulsion Laboratory, Pasadena, California, April 3, 1991.

tragalactic radio source is also used). It has also been suggested that postprocessed VLBI science data be used to improve the accuracy of the orbit [4], but this would only be possible for radio astronomy platforms and is not likely to benefit the orbit determination process unless real time, or at least, near-real-time correlation is possible. Other, more esoteric methods have been investigated, including the use of satellite laser ranging systems and onboard microaccelerometers. The advantages and disadvantages of these tracking methods have been discussed in greater detail in [5].

A data type that is geometrically equivalent to DOR, yet somewhat less cumbersome to implement operationally, is two-way minus three-way range (differenced range). Analogously, two-way minus three-way Doppler (differenced Doppler) can also be employed. These data types are usually referred to as "quasi-VLBI" data because they provide information about the spacecraft position and velocity similar to that provided by VLBI. Although less accuracy can be achieved than with VLBI, differenced Doppler or range data can be made available for navigational purposes faster than data from the DSN VLBI system. Until recently, differenced range was not thought to provide adequate accuracy for orbit determination due to large systematic errors associated with the data, in particular, clock offsets between co-observing stations. However, this method is becoming more attractive as better calibration systems become available. A recent analysis has shown that improvements in DSN ranging system calibration accuracies, together with clock offset and other calibration data derived from the GPS, could theoretically determine spacecraft angular coordinates to 30–90 nrad accuracy [6]. A similar investigation of the use of differenced Doppler for orbit determination, based on analysis of the 8.4-GHz (X-band) data acquired from the Magellan spacecraft, showed that differenced Doppler might deliver 50–100 nrad spacecraft angular accuracy, except for spacecraft within about 10 deg of the celestial equator [7]. Results from these analyses are encouraging enough to warrant investigation of these data types for use in HEO mission support.

This article presents an orbit determination accuracy assessment for an HEO mission set by using two-way range, differenced Doppler, and differenced range measurements together with conventional two-way Doppler. Three different HEO missions are studied, each of which is taken from the orbiting-VLBI (OVLBI) radio astronomy platform set: the Japanese VLBI Space Observatory Program (VSOP); the Soviet RADIOASTRON Project; and the International VLBI Satellite (IVS). Orbital characteristics for each mission are defined, with associated ground tracks and DSN view periods provided. An array of numerical error covariance analyses are performed to evaluate orbit determination accuracies achievable with different data strategies. A detailed description of the assumptions used for tracking data simulation and error modeling is also provided, along with a discussion of observable formulation and downlink footprint limitations, as well as a description of the orbit determination error-modeling parameters. In summary, a performance assessment is provided for each mission based on the numerical results generated during the study.

II. Mission Orbital Characteristics

A set of sample orbit parameters for VSOP and RADIOASTRON are summarized in Tables 1 and 2 (initial spacecraft ephemerides are referenced to Earth mean equator and equinox of date). Based on these orbital characteristics, each orbiter will trace repeatable ground tracks and VSOP will experience a nodal regression and apsidal rotation of roughly 180 deg per year due to Earth oblateness. The RADIOASTRON orbit will maintain a "fixed" perigee since its orbital inclination is taken to be near the critical value of 63.4 deg; the ground track required to remain constant with about a 24-hr period. A ground-track profile and DSN view periods for VSOP are shown in Figs. 1(a) and 1(b) for a typical 24-hr period.[2]

The sample RADIOASTRON orbit optimizes ground coverage over the Soviet Union, but has limited visibility from the DSN, with the exception of Madrid. To use differenced data types, at least two ground stations must simultaneously "view" the spacecraft for some interval of time, but the nominal orbital geometry does not permit a view period overlap for any DSN intercontinental baseline. Therefore, for this analysis, the longitude of the ascending node is shifted westward to 145 deg, which enables an overlap between Madrid and Goldstone. The corresponding ground track and DSN view periods for this shifted orbit are shown in Figs. 2(a) and 2(b).

In the case of the IVS, mission design remains in the early stages of development; a "nominal" orbit has yet to be clearly defined. There does exist an array of suggested orbital geometries ranging from a moderate-altitude (~25,000 km) to a high-altitude (~40,000 km) orbit, even an "ultra-high" altitude (~150,000 km) orbit has been suggested. A set of four orbit phases was selected in an ear-

[2] The labels DSCC 10, DSCC 40, and DSCC 60 in this and subsequent figures represent the Deep Space Communications Complexes (DSCCs) at Goldstone, California; near Canberra, Australia; and near Madrid, Spain, respectively.

lier IVS orbit determination study [8], but only the ultra-high altitude phase is considered here since the moderate- and high-altitude orbit cases are similar geometrically to VSOP and RADIOASTRON. The orbital parameters are summarized in Table 3, with the corresponding ground-track and DSN view periods shown in Figs. 3(a) and 3(b).

III. Assumptions for Numerical Error Covariance Analysis

A. Observable Development

The general model for the range observable used in the Orbit Analysis and Simulation Software (OASIS) package is defined by

$$\rho \equiv r + \tau_T + \tau_I + \tau_C \tag{1}$$

where

r = distance between the spacecraft and Earth-based tracking station[3]

τ_T = delay due to troposphere

τ_I = delay due to ionosphere

τ_C = delay due to clock offset calibration errors

A detailed mathematical description of observable models and their associated partial derivatives (partials) is provided in *OASIS Mathematical Description.*[4] The range rate, which is proportional to the Doppler observable, is determined by taking the time derivative of the range observable given in Eq. (1). The studies presented here for two-way measurements assume that appropriate linear combinations of the one-way measurements described by Eq. (1) can be formed for purposes of conducting error analyses. These assumptions apply to observable formulation and computation of the partials; moreover, round-trip light time corrections are omitted in the formulation of the observations and partials (even though this capability does exist) since these corrections are normally reserved for interplanetary studies. Partials for ionospheric refraction are omitted in this analysis since it is expected that the sensitivity due to tropospheric path delay will be the major propagation media effect due to the high operating frequencies proposed for OVLBI (i.e., 8–15 GHz).

[4] S. C. Wu, W. I. Bertiger, J. S. Border, S. M. Lichten, R. F. Sunseri, B. G. Williams, P. J. Wolff, and J. T. Wu, *OASIS Mathematical Description, V. 10*, JPL D-3139 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 1, 1986.

Constructing an observable for differenced Doppler/range amounts to differencing a two-way and a three-way measurement. The resulting observable is a by-product of differencing the downlink signals of each participating Earth-based tracking station over a common view period (see Fig. 4). Mathematically, this can be expressed as

$$\Delta\dot{\rho}(\Delta\rho) = \underbrace{\dot{\rho}_{1u}(\rho_{1u}) + \dot{\rho}_{1d}(\rho_{1d})}_{\text{2-way measurement}} - \underbrace{[\dot{\rho}_{1u}(\rho_{1u}) + \dot{\rho}_{2d}(\rho_{2d})]}_{\text{3-way measurement}} \tag{2}$$

$$= \dot{\rho}_{1d}(\rho_{1d}) - \dot{\rho}_{2d}(\rho_{2d})$$

where

$\dot{\rho}_{1u}(\rho_{1u})$ = uplink Doppler/range measurement from DSN station 1 to the spacecraft

$\dot{\rho}_{1d}(\rho_{1d})$ = downlink Doppler/range measurement from the spacecraft to DSN station 1

$\dot{\rho}_{2d}(\rho_{2d})$ = downlink Doppler/range measurement from the spacecraft to DSN station 2

B. Tracking Data Simulation

For the error analyses, two-way 8.4-GHz (X-band) Doppler data were taken to be the primary data type and were collected continuously from two Deep Space Stations (DSSs), at different intervals, throughout the course of a single orbit arc.[5] The two stations were chosen so that a region of overlap existed in which differenced Doppler and range measurements could also be acquired.

To account for data noise, the Doppler was weighted with a 1-σ measurement uncertainty of 0.1 mm/sec over a 60-sec integration time. The data weight ($1/\sigma^2$) was adjusted according to the minimum elevation at the station, and data points below 10-deg local horizon were omitted. In certain cases, two-way 2.3-GHz (S-band) range data were acquired continuously along with Doppler but sampled at a rate of one point every 240 seconds. These data were also weighted for minimum elevation, and the 1-σ measurement uncertainty was taken to be 15 cm. As with Doppler, range data were collected only when the spacecraft was considered in-view—when the station elevation angle is greater than 10 deg.

Differenced Doppler and range measurements were collected (assuming a three-way link) only in the near-

[5] It is expected that the VSOP mission will use a 15-GHz link; the analysis in this article is meant as an illustrative case using the VSOP orbit, not as a specific application for the mission.

apoapsis region of the orbit. Differenced Doppler measurements were sampled over a 60-sec integration time with elevation corrections. The 1-σ measurement uncertainty was assumed to be 0.15 mm/sec—50 percent more noisy than two-way Doppler data. Differenced range measurements were weighted (also according to minimum elevation) with 1-σ measurement uncertainty of 22.5 cm, and with a data point acquired every 2 minutes.[6]

Four data strategies were assumed for each OVLBI mission: two-way Doppler only, two-way Doppler plus two-way range, two-way Doppler plus differenced Doppler, and two-way Doppler plus differenced range.

An important caveat to note is that it may not be possible to construct differenced Doppler/range measurements due to the limited ground signature (footprint) of the parabolic downlink antenna designs expected for the OVLBI. This could be a major obstacle if these data types are sought to be employed operationally, especially for the large intercontinental baselines of the DSN and with high link frequencies. Table 4 shows a small sample of antenna footprints (computed from beamwidth = frequency/diameter) that assume a parabolic downlink antenna operating at 8.4 GHz (X-band). (Note that the Goldstone–Madrid baseline is roughly 8,600 km in length, while the Goldstone–Canberra baseline is about 10,600 km.)

It is clear from the table that the expected slant ranges for the VSOP are not nearly large enough to facilitate the acquisition of three-way data for any DSN intercontinental baseline; not to mention the fact that the 15-GHz link further reduces the size of the ground footprint. For RADIOASTRON, it may be possible to acquire three-way data from a Goldstone–Madrid baseline if a small (~0.25-m diameter) downlink antenna is employed and if data are collected near the apoapsis regions of the orbit. Of the three orbits considered in this study, the ultra-high altitude IVS orbit clearly would be most suitable for the acquisition of three-way data. Perhaps for future HEO missions, a variety of antenna-link designs and configurations (e.g., multiple, independently pointable antennas) will be considered.

[6] These may be slightly more "conservative" estimates of the measurement uncertainty; if independent and random measurements were assumed for the differenced data types investigated here, one could argue that the measurement uncertainty for differenced Doppler/range could be determined $\sigma_{\Delta\dot{\rho}}(\sigma_{\Delta\rho}) = \sqrt{2}_{\sigma\dot{\rho}}$ $(\sqrt{2}\sigma\rho)$. In this case, the one-sigma uncertainties for differenced Doppler/range would be approximately 0.14 mm/sec and 21.2 cm, respectively.

C. Orbit Determination Modeling Errors[7]

The modeling of error sources was performed in essentially the same manner for all three OVLBI missions and was based on the current and expected future performance of the spacecraft and the DSN. The fundamental error-modeling assumptions used in this analysis were based on Konopliv's earlier work,[8] but were somewhat modified and expanded for completeness.

The error-modeling parameters are broken down into two categories: estimated and considered. The random data noise characteristics addressed in Section III.B are summarized in Table 5, along with the estimated parameters, considered parameters, and associated sigmas.

1. Estimated Parameters. The estimated parameters were chosen to account for mismodeling of spacecraft nongravitational accelerations and calibration errors for certain data types. Random accelerations due to solar radiation pressure (SRP) and gas leaks were treated as stochastic processes and were estimated along with the spacecraft trajectory.

For the VSOP, RADIOASTRON, and the IVS spacecraft, the area-to-mass ratios were taken to be 0.04, 0.03, and 0.10 m^2kg^{-1}, respectively. (The small area-to-mass ratio for VSOP occurs because the onboard VLBI antenna is expected to be nearly 80 percent transmissive, whereas the the small ratio for RADIOASTRON results from the large mass, 4,000 kg, of the spacecraft). The SRP specular and diffuse reflectivity coefficients were designated stochastic variables, and the stochastic model for process noise was taken to be a first-order Markov colored noise model (exponentially correlated process) with steady state sigmas equivalent to 10 percent of the maximum attainable a priori values. Time constants were chosen to be roughly equivalent to the orbital period in each case (1/4 day for the VSOP, 1 day for RADIOASTRON, and 3 days for the IVS). A batch-sequential, factorized Kalman filter was used in the estimation process, with batch sizes for each of the aforementioned missions of 1 hr, 3 hr, and 6 hr, respectively.

Gas leak accelerations (in each spacecraft body-fixed axis) were estimated stochastically with steady state sig-

[7] A quick semantic note: The words *error* and *uncertainty* are used interchangeably throughout the text and, as such, they are intended to be synonymous; moreover, an error, in this context, is intended to imply uncertainty in measurement, not a mistake or blunder.

[8] A. Konopliv, "Preliminary Orbit Determination Analysis for the VSOP Mission," JPL Interoffice Memoranda 314.4-648 and 314.4-667 (internal documents), Jet Propulsion Laboratory, Pasadena, California, February 9 and July 20, 1989.

mas of 10^{-12} km/sec^2, and time constants were assumed to be 1 day for the VSOP and RADIOASTRON and 3 days for the IVS. The batch sizes were taken to be equivalent to the SRP batch sizes.

To account for the effects of station-frequency offset-calibration errors on differenced Doppler data, a bias parameter was added to these measurements. The bias parameter was estimated as a constant parameter with 1-σ uncertainty of the total a priori offset calibration value of 0.02 mm/sec. In the case of two-way range, a bias parameter was added for each DSS receiver to account for clock offsets and signal path delays. Similarly, for differenced range, a clock offset or signal path delay bias parameter was added. The 1-σ uncertainty of the total a priori offset calibration value was taken to be 300 cm (~20 nsec). These uncertainties are representative of present-day capabilities. In the future, it is expected that GPS-based calibration methods will substantially reduce these uncertainties.

2. Considered Parameters. A *consider* parameter is treated by the filter as an unmodeled systematic error and may significantly affect the error statistics of the estimated parameter set. The total error covariance accounts for the consider variances as well as the variances computed by the filter, so as not to understate the predicted navigational performance. The considered parameters used in this study accounted for systematic errors in station locations, offset in the geocenter, gravity modeling, and tropospheric path delay.

Station uncertainties include both a relative component and an absolute (geocentric) component. The relative component refers to DSN-site–DSN-site uncertainty (measured accurately by VLBI); the geocentric component refers to a common error in locating the DSN sites with respect to the Earth mass center (VLBI is insensitive to this component). Conservative equatorial station location errors of 50–75 cm were assumed;[9] the relative error (station–station) being about 30 cm. For the z-direction, 10-cm relative and 1-m geocentric errors were assumed. Analysis indicates that the expected accuracy of the geocenter can be determined to better than 10 cm by using GPS-based measurements [9]. The majority of this analysis assumes that no such calibrations were available, again, for conservatism; the only exception being one special case in which GPS-based measurements were indeed assumed for calibrating the limiting error sources associated with each data type. This special case is further described in the next section.

Gravity-error modeling included the Earth's Newtonian gravitational parameter (GM) and an 8×8 reduced-order GEM-T2 gravity field obtained from NASA–Goddard Space Flight Center's 50×50 field [10]. Uncertainty of the GM was taken to be 1 part in 10^8, while the formal sigmas for the GEM-T2 gravity field harmonics were used and assumed to be uncorrelated.[10]

The wet and dry components of error contribution due to tropospheric path delay were considered with zenith uncertainties of 4 cm and 1 cm, respectively (based on present-day values from a seasonal model).

Atmospheric drag and Earth albedo (radiation pressure reflected from Earth) were omitted from the error-modeling process because their effects were shown to contribute less than 1 cm to the total orbit determination error in an earlier analysis [8].

IV. Performance Assessment

The orbit determination accuracy results described here are expressed in terms of uncertainty in spacecraft position and velocity. All stated results represent root-mean-square (rms) or 1-σ performance statistics, since the batch-sequential filter in this analysis used linear unbiased estimation methods. The filter-generated computed covariance was combined with consider parameter sensitivities to construct the total or *full consider covariance*, which was then mapped forward in time to produce a time history of propagating error sources and navigational performance for four data strategies (Doppler only, Doppler plus two-way range, Doppler plus differenced Doppler, and Doppler plus differenced range). The covariance was mapped ahead for a duration of 24 hr in the case of the VSOP and RADIOASTRON orbits, and 72 hr in the case of the IVS orbit.

A. VSOP Orbit

The accuracy statistics for VSOP position and velocity uncertainties are shown in Figs. 5(a) and 5(b). These figures represent the total uncertainty over time, in a root-sum-square (rss) sense, of all vector components,

[9] T. Moyer, "Station Location Sets Referred to the Radio Frame," JPL Interoffice Memorandum 314.5-1334 (internal document), Jet Propulsion Laboratory, Pasadena, California, February 24, 1989.

[10] Another, and perhaps more common, approach to gravity field mismodeling is to assume a "lumped sum" uncertainty of a fraction (say 50 percent) of two *differenced* gravity fields (e.g., GEM-10 and GEM-L2). This is usually the method of choice when computational disk space and processing time are at a premium.

i.e., radial (altitude), transverse (down-track), and normal (cross-track), which account for the computed plus consider errors.

Two-way Doppler and range were collected only during the first passes at Madrid (00:17 to 03:57 hr) and Goldstone (06:16 to 09:04 hr), a little over 7 hr in all; no other stations were used for data acquisition. To construct differenced Doppler and differenced range measurements, a 90-min three-way link between 07:34 and 09:04 hr was assumed for a Madrid–Goldstone baseline (baseline slant ranges: ~18,000–24,000 km). The performance of the Doppler-only and Doppler-plus-range solutions was very similar throughout the reconstruction and prediction phases of the propagation cycle. By augmenting the Doppler-only data set with a relatively short pass of differenced Doppler or differenced range data, a dramatic improvement in orbit determination performance resulted, which suggests that, by using differenced Doppler data over the data arc, there is about a 59-percent improvement in position performance and about a 78-percent improvement in velocity performance. When differenced range data are used, there is about a 43-percent improvement in position determination and about a 63-percent improvement in velocity performance.

Figure 6 represents a "snapshot" of performance statistics for position and velocity taken at the initial apogee crossing and at the first perigee passage, respectively. The total uncertainty is broken down into the orthogonal components of radial, transverse, and normal error. Again, the Doppler-only and Doppler-plus-range measurements yield similar characteristics, while the differenced measurements improve results more dramatically. It is especially interesting to note that the Doppler plus differenced-Doppler measurements were significantly better able to determine the out-of-plane (normal) component than the other data strategies investigated.

For the same snapshot event, a breakdown of individual error sources is provided (see Fig. 7). The computed statistics represent the estimated or formal filter results and are frequently referred to as data noise contributions. The differenced Doppler measurements appear to yield the best performance with respect to each error source. The important observation, however, is that, by augmenting the two-way Doppler data with either differenced Doppler or differenced range data, the effect of the tropospheric uncertainty, which is the dominant error source here, is substantially reduced. This behavior occurs because line-of-sight data types, such as two-way Doppler and range, are very sensitive to tropospheric zenith-delay calibration errors, whereas differenced Doppler and range data types are relatively less sensitive to the calibration errors. Another important observation is the power of the differenced data types for reducing velocity uncertainty in the near-perigee regime, again, with the tropospheric uncertainty contribution being dramatically reduced.

B. RADIOASTRON Orbit

In some respects, the covariance results for the RADIOASTRON position uncertainty resemble those of the VSOP, despite different orbital characteristics. However, this is not true in the case of velocity uncertainty. Results for velocity performance are omitted due to the less than 10-percent difference among all four data strategies investigated (see Table 6 for a summary of all results). Accuracy statistics for position uncertainty are provided and summarized in Fig. 8.

Two-way Doppler and range data were collected from Goldstone and Madrid between 00:34 and 13:43 hr and 14:00 to 23:34 hr, respectively. This amounted to nearly a 23-hr data arc—substantial coverage, indeed. Differenced Doppler and differenced range measurements were constructed for a 4-hr Goldstone–Madrid baseline between 09:42 and 13:42 hr (baseline slant ranges: ~71,000–74,000 km). The results indicate about a 6-percent improvement in position determination when range data are included with the Doppler; this is clearly evident in the region ±6 hr of apoapsis. The augmented differenced Doppler case yields about a 7-percent improvement, but begins to degrade late in the day. Measuring Doppler plus differenced range data appears to be the best strategy in terms of orbit determination performance throughout the entire data arc. Results indicate about a 13-percent improvement over Doppler-only solutions.

Performance statistics taken at the initial apogee crossing in terms of orthogonal components and individual error sources are shown in Figs. 8(b) and 8(c), respectively. Here, all four data strategies yield similar performances in terms of the radial and transverse components, with the exception of the normal component, which was better determined by using differenced Doppler or differenced range data.

From the error-breakdown chart, Fig. 8(c), the dominant error sources appear to be data noise, station locations, and troposphere. The differenced data strategies, particularly differenced range, seem to reduce the uncertainty in the troposphere, whereas for the VSOP orbit, the differenced Doppler helped to reduce the uncertainty in the troposphere better than the other data types. This phenomenon reflects the fact that there is greater information content in the differenced Doppler measurements

for the lower altitude, highly elliptical orbiters, since the data are very sensitive to the spacecraft's orbital motion. The differenced range information content is greater for the higher altitude elliptical orbiters because the data are more sensitive to orbital geometry than dynamics.

C. IVS Orbit

The IVS ultra-high altitude orbit case (67-hr orbit) offers an interesting problem in terms of scheduling tracking passes. For this study, the argument of perigee and longitude of ascending node were "optimized" to provide substantial data coverage, particularly for the northern hemisphere stations. As in the RADIOASTRON cases, the covariance results for velocity uncertainty amounted to less than a 15-percent difference for all four data strategies, which translates to only ~0.13 mm/sec improvement at best; therefore, these statistics are not shown. The position uncertainty is certainly dependent upon the assumed data strategy, and accuracy statistics are provided in Fig. 9.

Several long passes of two-way Doppler and range data can be obtained from Madrid and Goldstone for this orbit. The tracking schedule is taken to be: 00:38 to 14:46 hr (Madrid) and 15:00 to 23:06 hr (Goldstone) for the first day; 04:56 to 01:44 hr (Goldstone) for the second-to-third day; and 02:00 to 17:48 hr (Madrid) for the third day, a total of ~59 hr. Ten hours of three-way data were acquired for a Madrid–Goldstone baseline ±5 hr either side of apogee (baseline slant ranges: ~148,000–151,000 km). The results show that the Doppler-plus-range strategy yields the best performance during the early part of the orbit arc (i.e., for the first day), but then begins to degrade at the start of the second day. For this case, about a 13-percent improvement in performance resulted between $P_0 + 6$ hours and $P_1 - 6$ hours, where P_0 represents the initial perigee point (at epoch) and P_1 represents the first perigee passage (end of the orbit arc). When differenced Doppler and differenced range data are used, better performance clearly results throughout most of the reconstruction phase, with differenced range data yielding the best results, about a 20-percent improvement in performance, while differenced Doppler provided only about a 14-percent improvement.

Figures 9(b) and 9(c) give the accuracy statistics for all four data strategies in terms of orthogonal components and error characteristics. As with the RADIOASTRON cases, the radial component is determined about equally as well for the four data strategies, and only slightly better for Doppler-plus-range data. However, unlike RADIOASTRON, the transverse component is better determined with the augmented data. The normal component is, again, best determined with the differenced data types.

Limiting error sources for this case are seen to be data noise and station location uncertainties, with the data noise component for the two-way Doppler-only case dominating all other error sources. This should not be surprising because of the long data arcs used in this case. Effects of gravity mismodeling are, as expected, negligible because of the high-altitude phase of this orbit (~150,000 km). Unlike the VSOP and RADIOASTRON, the data strategy of Doppler plus range appears to be most sensitive to tropospheric calibration errors.

To gain insight into what orbit determination accuracies can be achieved by using these radiometric data types together with GPS-based ground observations for measuring and calibrating the major systematic error sources, a second set of numerical covariance analysis runs was made for IVS. Results are shown in Fig. 10, again with velocity omitted because of the relatively insignificant difference in performance among data strategies (23-percent improvement at best, which translates to ~0.16 mm/sec).

For this study, a tighter error budget was assumed with respect to the observing platform and propagation media errors, as well as the biases associated with each data type. Specifically, relative station location uncertainties were reduced to 7 cm (each component), the wet zenith troposphere delay to 2 cm, and the geocenter offset to 10 cm. These values are still considered to be rather conservative, especially for a mid-to-late 1990s time frame.[11] The a priori frequency bias was then taken to be 0.003 mm/sec (1 part in 10^{14}), while a priori biases due to clock offset and signal path delays were taken to be 2 nsec (60 cm).

The same tracking passes were assumed in the earlier case. The Doppler-only solution is improved over the original Doppler-only solution, and the augmented data strategies also yield better performance than in the original case: about 29-percent improvement for Doppler plus range data, 32-percent improvement for Doppler plus differenced-Doppler data, and 39-percent improvement for Doppler plus differenced-range data.

In terms of orthogonal components, the radial error component is again about the same for all strategies, but transverse and normal components were significantly better determined by using the augmented data types. Clearly, when used in concert with the augmented data

[11] S. Lichten, private communication, Tracking Systems and Application Section, Jet Propulsion Laboratory, Pasadena, California.

types assumed in this analysis, the GPS-based ground observations reduced the major systematic error sources substantially enough, see Fig. 10(c), that performance becomes limited only by the quality of the data. Similar performance improvements were evident in earlier orbit determination analyses of the actual VSOP and RADIOASTRON missions.[12,13]

These results are very encouraging because they suggest that radiometric data types, such as two-way Doppler, can deliver precise, orbit determination accuracies when augmented with differenced data types and GPS-based ground-calibration data. It is hoped that such a ground-calibration system will become operational not only for support of future HEO missions, but also for deep-space missions.

Table 6 gives a summary chart of orbit determination accuracy results (rss total uncertainty in position and velocity) for all the different data strategies and missions investigated. A range of statistics is displayed which indicates the rms or 1-σ orbit determination accuracy results accumulated over the reconstruction phases (data arcs) of a single orbit arc for each mission. For this analysis, the relative improvement is difficult to quantify because it varies over different phases of the orbit, and knowledge at some points may be more important than at others. Table 6 attempts to quantify the improvement in two ways, giving both the range in uncertainties and the average (percentage) improvement over the data arcs. The actual values used to produce the percentages of improvement were computed by integrating each error curve over the data arc to obtain the total area relative to the total area of the reference data strategy of two-way Doppler only.

V. Remarks

In this analysis, long two-way Doppler passes were used, which amounted to obtaining all possible data available from two ground telemetry sites. This, of course, may not be realizable in an operational setting without a dedicated ground network of antennas. Nevertheless, by utilizing all the Doppler data available and comparing the resulting performance with that obtained from relatively short passes of differenced data, one can ascertain the true power of the differenced data types. The results suggest that the more two-way data available, the less the influence of the differenced data on the resulting orbit determination accuracy. However, it is not yet clear if the improvement is strictly a function of the orbital motion/geometry or due to the relative lengths of the data arcs; most likely, it is a combination of both (recall that the VSOP orbit determination results using differenced Doppler/range yielded the most dramatic results). This makes intuitive sense since the VSOP maintained a lower altitude than its counterparts, RADIOASTRON and IVS, and the lower the altitude, the greater the parallax between the co-observing stations; hence, information content of the data is increased. It could be that if substantial coverage (i.e., two-way Doppler data) cannot be obtained due to other antenna demands, use of even a limited amount of differenced data would provide increased orbit determination accuracy. The availability of GPS calibrations may assume greater significance if the data quantities are limited.

It is also important to note that only east–west baselines were investigated for this study. This was not intentional, but was a result of the fact that the orbital characteristics of the cases examined were simply better suited for east–west baselines. It is not yet clear whether better performance can be achieved with north–south baselines.

The downlink antenna footprint limitation can be of serious consequence operationally, especially if high link frequencies (~8 GHz and higher) are to be used. Therefore, future missions may require the use of two downlink antennas to obtain the possible performance improvements seen here.

VI. Conclusions

An orbit determination analysis was performed for a set of three different orbits being considered as orbiting radio astronomy (OVLBI) platforms to investigate the utility of differenced Doppler and range measurements. The orbital characteristics were described along with a discussion of the assumptions used for numerical error covariance calculations. The covariance analyses were performed by using a factorized Kalman filter to estimate a set of governing parameters and to compute the sensitivity of the estimated parameters to a set of unmodeled considered parameters that were treated as systematic error sources. Orbit determination accuracy statistics were calculated for all three orbits with an additional case reflecting the use of GPS-based ground observations for calibrating all radiometric data types investigated (two-way Doppler/range and differenced Doppler/range).

[12] C. S. Christensen and J. A. Estefan, "Orbit Determination Capability for VSOP Using DSN Doppler Tracking," JPL Interoffice Memorandum 314.5-1424 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 15, 1990.

[13] C. S. Christensen and J. A. Estefan, "Orbit Determination Capability for Radioastron Using DSN Doppler Tracking," JPL Interoffice Memorandum 314.5-1448 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 4, 1990.

Results indicate that for the VSOP orbit, an accuracy improvement of up to 59 percent was achieved by using two-way Doppler plus differenced Doppler for position determination, and up to 78-percent improvement in velocity determination by using the same data strategy. These were, by far, the most dramatic results seen in the analysis. For the RADIOASTRON and the IVS orbits, the relative improvement in velocity by using two-way range measurements or even differenced measurements was rather insignificant. However, up to 20-percent improvement in position determination was evident when differenced range measurements were used to augment the two-way Doppler measurements. The power of the differenced data types was even more evident when GPS-based ground calibrations were assumed to measure and calibrate the limiting systematic error sources. In the case of the IVS orbit, orbit determination accuracy improved by roughly 39 percent in position and 23 percent in velocity.

In exploring potential tracking methods for highly precise (submeter) orbit determination, earlier orbit determination analyses [8] focused on exploiting conventional two-way Doppler tracking together with GPS-based tracking methods, which included ground-based calibration data as well as onboard flight receivers. The results suggested that the two tracking techniques complemented each other extremely well throughout the orbit arcs, and it would be interesting to investigate what accuracies might be achieved when differenced Doppler and differenced range measurements are used in the tracking processes as well.

Acknowledgments

The author is indebted to Jim Ulvestad for providing the essential material regarding downlink footprint limitations and for his critical review of the draft report. His comments and suggestions proved invaluable in completing this work effort. A warm thanks to Sam Thurman for inspiring this work effort; and to him, Carl Christensen, and Steve Lichten for their careful review of the subject matter.

References

[1] R. B. Frauenholz and J. Ellis, "Orbit Determination of Highly Elliptical Earth Orbiters Using VLBI and ΔVLBI Measurements," *TDA Progress Report 42-75*, vol. July–September 1983, Jet Propulsion Laboratory, Pasadena, California, pp. 1–13, November 15, 1983.

[2] T. P. Yunck, W. G. Melbourne, and C. L. Thornton, "GPS-Based Satellite Tracking System for Precise Positioning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-23, pp. 450–457, July 1985.

[3] S. Gorgolewski, "Precise RADIOASTRON Baseline Determination," paper presented at the discussion meeting on the "RADIOASTRON project: Navigation, Astrometry, Geodesy Aspects" of the SGO-IKI Joint Working Group, Satellite Geodetic Observatory, Penc, Hungary, February 27–March 3, 1989.

[4] I. Fejes, T. Borza, Sz. Mihály, and L. Szánthó, "Orbit Determination Accuracy Improvement by Space-VLBI Observables as Tracking Data," paper presented at the IAG General Meeting, Symposium 101: Global and Regional Geodynamics, Edinburgh, Scotland, August 3–5, 1989.

[5] I. Fejes, ed., *Precise Orbit Determination of RADIOASTRON*, report prepared by the RADIOASTRON Navigation, Astrometry, and Geodesy (NAG) Group for the 8th RADIOASTRON Review meeting, ver. 29, Satellite Geodetic Observatory, Penc, Hungary, April 1989.

[6] S. W. Thurman, "Deep-Space Navigation with Differenced Data Types, Part I: Differenced Range Information Content," *TDA Progress Report 42-103*, vol. July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 47–60, November 15, 1990.

[7] S. W. Thurman, "Deep-Space Navigation with Differenced Data Types, Part II: Differenced Doppler Information Content," *TDA Progress Report 42-103*, vol. July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 61–69, November 15, 1990.

[8] S. M. Lichten and J. A. Estefan, *High Precision Orbit Determination for High-Earth Elliptical Orbiters Using the Global Positioning System*, paper AIAA 90-2954, AIAA/AAS Astrodynamics Conference, Portland, Oregon, August 20–22, 1990.

[9] R. P. Malla and S. C. Wu, "The Geocenter Estimation Results Using GPS Measurements," *GPS '90* (in press), September 3–7, 1990, paper presented at the Second International Symposium on Precise Positioning with the Global Positioning System, Ottawa, Canada, September 1990.

[10] J. G. Marsh, F. J. Lerch, B. H. Putney, T. L. Felsentreger, B. V. Sanchez, S. M. Klosko, G. B. Patel, J. W. Robbins, R. G. Williamson, T. L. Engelis, W. F. Eddy, N. L. Chandler, D. S. Chinn, S. Kapoor, K. E. Rachlin, L. E. Braatz, and E. C. Pavlis, "The GEM-T2 Gravitational Model," *Journal of Geophysical Research*, vol. 95, no. B13, pp. 22,043–22,071, December 10, 1990.

Table 1. Sample orbit parameters for the Japanese Very Long Baseline Interferometry Space Observatory

Parameter	Value
Epoch	
Date	January 1, 1995
Time	$00^h:00^m:00^s.0000$ UTC
Initial spacecraft ephemeris	
Semi-major axis	16,878 km
Eccentricity	0.5629
Inclination	46.0°
Argument of perigee	0.0°
Longitude of ascending node	0.0°
Mean anomaly	0.0°
Additional parameters	
Perigee height	1,000 km
Apogee height	20,000 km
Orbit period	6.06 hr

Table 2. Sample orbit parameters for the Soviet RADIOASTRON Project

Parameter	Value
Epoch	
Date	August 1, 1993
Time	$00^h:00^m:00^s.0000$ UTC
Initial spacecraft ephemeris	
Semi-major axis	42,378 km
Eccentricity	0.802303
Inclination	65.0°
Argument of perigee	285.0°
Longitude of ascending node	200.0°
Mean anomaly	0.0°
Additional parameters	
Perigee height	2,000 km
Apogee height	70,000 km
Orbit period	24.10 hr

Table 3. Orbit parameters for ultra-high altitude International VLBI Satellite Orbit

Parameter	Value
Epoch	
Date	January 1, 1999
Time	$00^h:00^m:00^s.0000$ UTC
Initial spacecraft ephemeris	
Semi-major axis	83,878 km
Eccentricity	0.86435
Inclination	63.0°
Argument of perigee	270.0°
Longitude of ascending node	120.0°
Mean anomaly	0.0°
Additional parameters	
Perigee height	11,378 km
Apogee height	156,378 km
Orbit period	67.14 hr

Table 4. Approximate parabolic antenna footprint, km, for 8.4-GHz (X-band) downlink

Slant Range, km	Antenna diameter, m		
	0.25	0.50	0.75
60,000	8,571	4,286	2,857
80,000	11,429	5,714	3,810
100,000	14,286	7,143	4,762
120,000	17,143	8,571	5,714
140,000	20,000	10,000	6,667
160,000	22,857	11,429	7,619

Table 5. Orbit determination error model parameters

Parameter	Uncertainty, 1σ
Random data noise	
Two-way Doppler	0.1 mm/sec
Two-way range	15 cm
Differenced Doppler	0.15 mm/sec
Differenced range	22.5 cm
Estimated parameters (a priori)	
Spacecraft	
Position	1,000 m
Velocity	1 km/sec
Stochastic accelerations	
Solar radiation pressure	10% of max
Gas leaks	10^{-9} m/sec^2
Biases	
Frequency offset	0.02 mm/sec
Clock offset/signal path delay	300 cm (20 nsec)
Considered parameters	
Gravity	
Earth's GM	GM$\times 10^{-8}$
Harmonics	8×8 field (GEM-T2)
DSN station coordinates	
Crust fixed	
Spin radius and longitude	50–75 cm
Z-height	10 cm
Geocenter-Z	1 m
Zenith troposphere delay	
Wet	4 cm
Dry	1 cm

Table 6. 1-σ orbit determination accuracy comparison for different data strategies

Data Strategy	RSS Position Uncertainty, m, and Relative Percentage Improvement			
	VSOP	RADIOASTRON	IVS	IVS + GPS Calibrations
Doppler-only versus	6.0–18.1	2.8–11.8	3.1–13.5	2.3–10.6
Doppler + range	5.6–16.8 6%	2.5–11.0 6%	2.8–11.9 13%	1.8–7.5 29%
Doppler + Δ Doppler	2.3–7.0 59%	2.5–11.0 7%	3.0–11.8 14%	2.2–8.4 32%
Doppler + Δ range	3.1–9.9 43%	2.4–10.3 13%	3.0–11.5 20%	2.2–8.2 39%
Data Strategy	RSS Velocity Uncertainty, cm/sec, and Relative Percentage Improvement			
	VSOP	RADIOASTRON	IVS	IVS + GPS Calibrations
Doppler-only versus	0.18–0.85	0.030–0.33	0.020–0.24	0.015–0.21
Doppler + range	0.17–0.78 7%	0.030–0.30 6%	0.016–0.21 13%	0.011–0.17 22%
Doppler + Δ Doppler	0.04–0.18 78%	0.030–0.32 5%	0.015–0.24 10%	0.011–0.21 20%
Doppler + Δ range	0.06–0.24 63%	0.028–0.31 10%	0.014–0.23 15%	0.010–0.20 23%

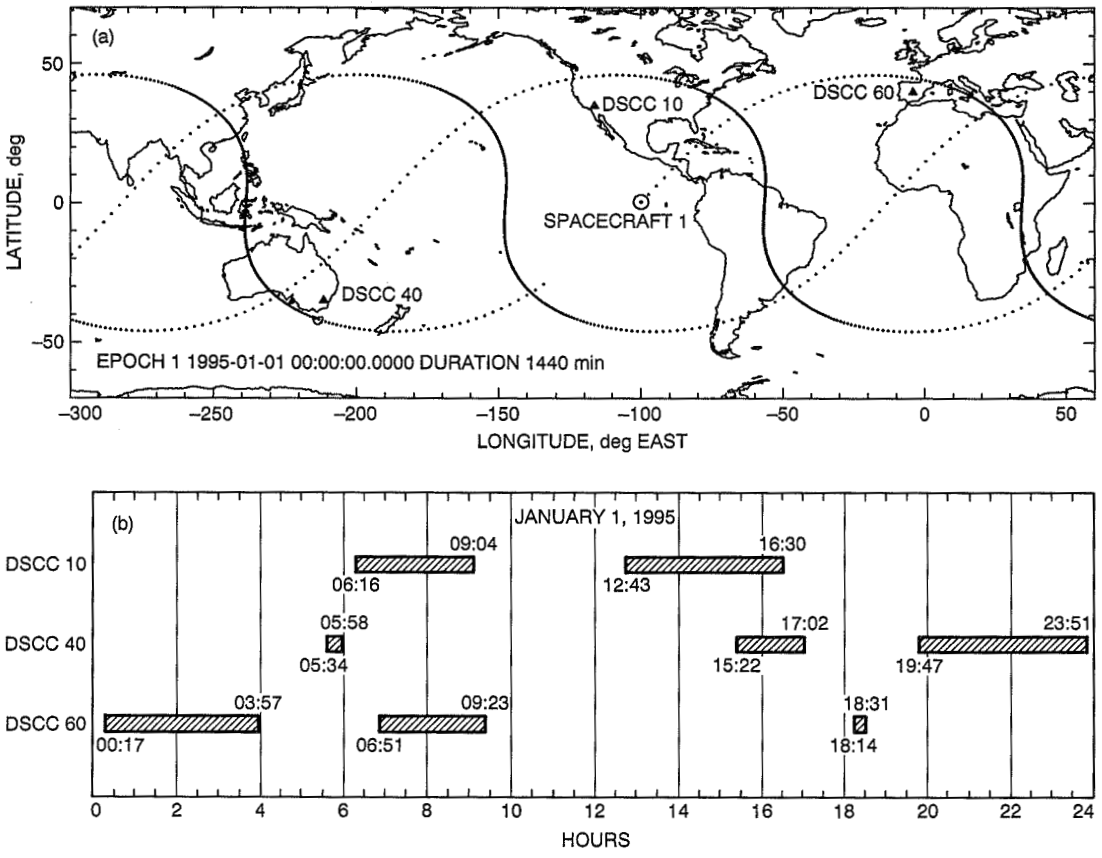


Fig. 1. DSN visibility profile for the VSOP sample orbit: (a) spacecraft ground track (1-min spacing); (b) ground network view periods (spacecraft rise/set times).

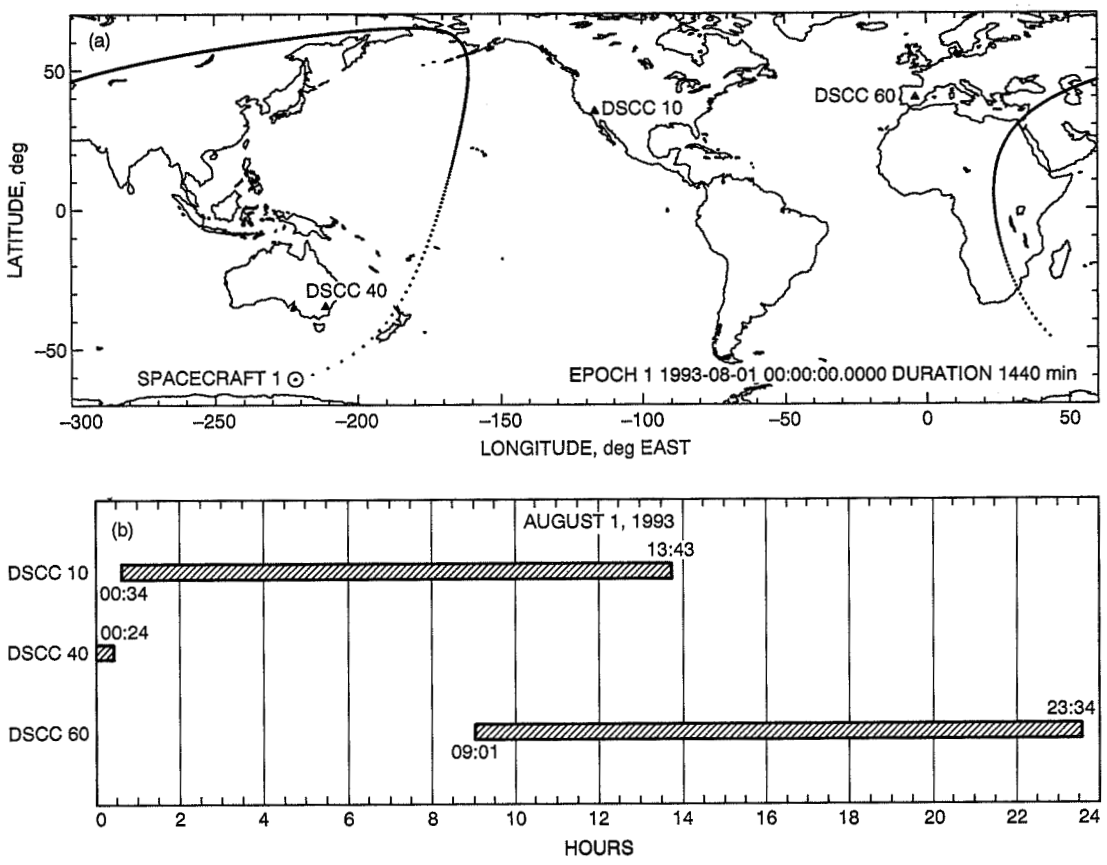


Fig. 2. DSN visibility profile for the RADIOASTRON shifted orbit: (a) spacecraft ground track (1-min spacing); (b) ground network view periods (spacecraft rise/set times).

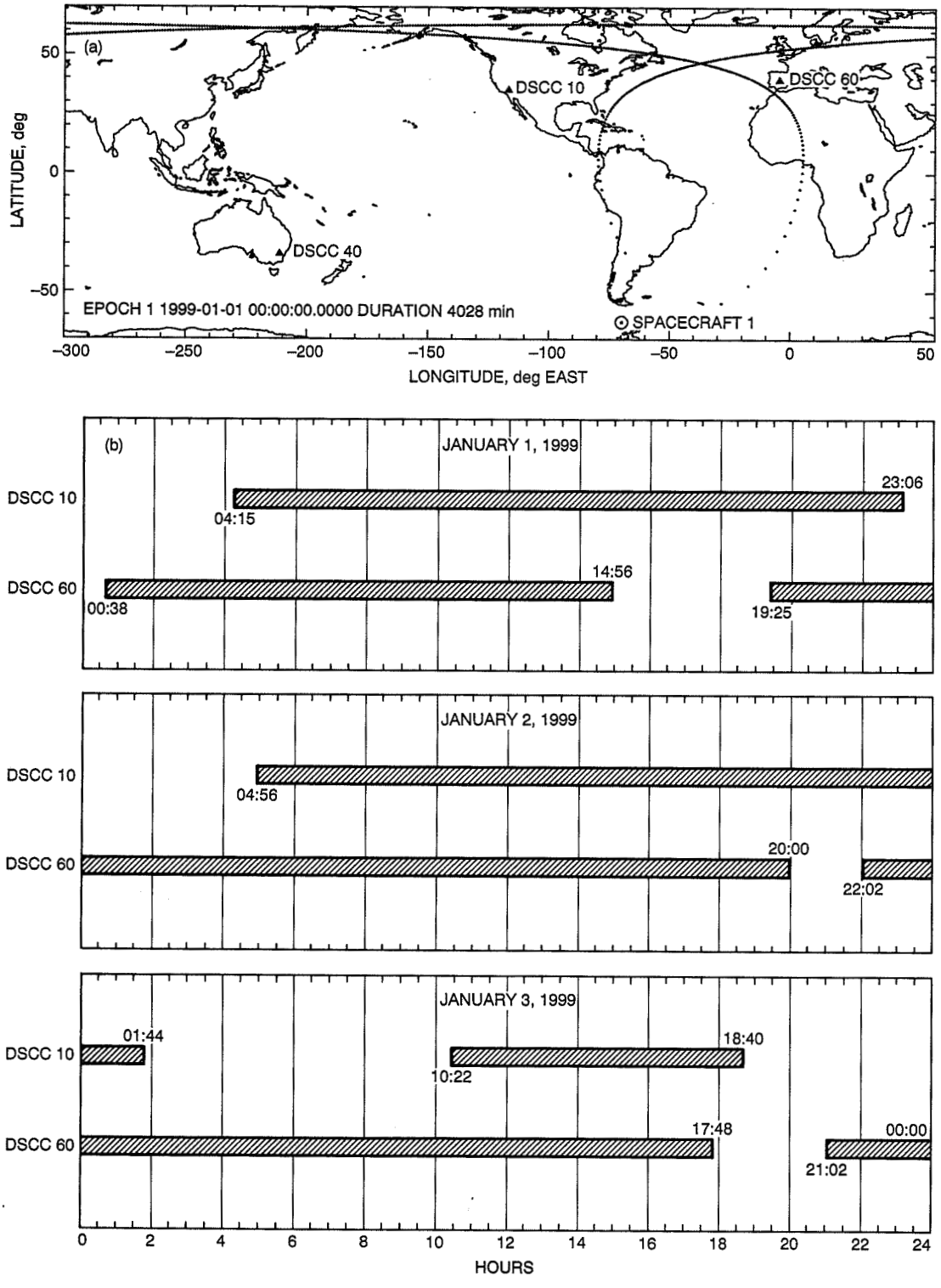


Fig. 3. DSN visibility profile for the ultra-high altitude IVS orbit: (a) spacecraft ground track (5-min spacing); (b) ground network view periods (spacecraft rise/set times).

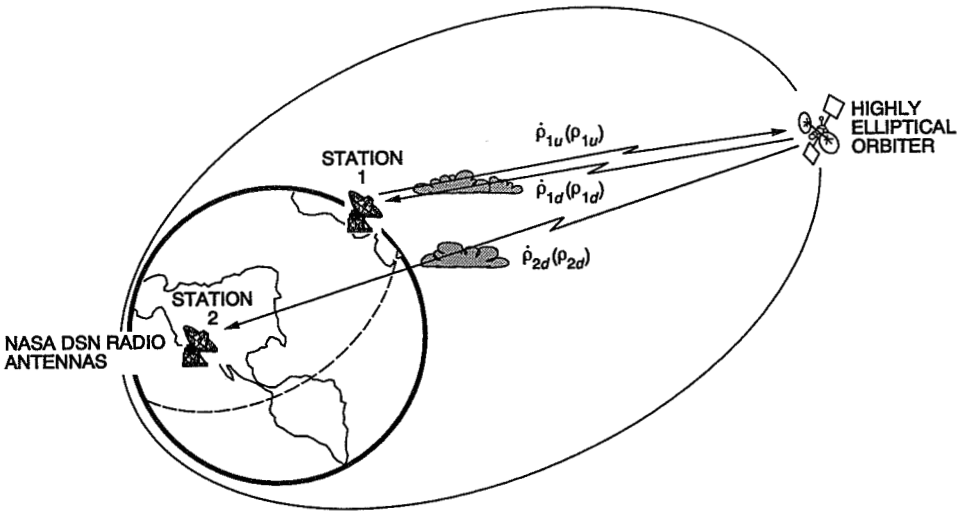


Fig. 4. Differenced Doppler/range acquisition scheme.

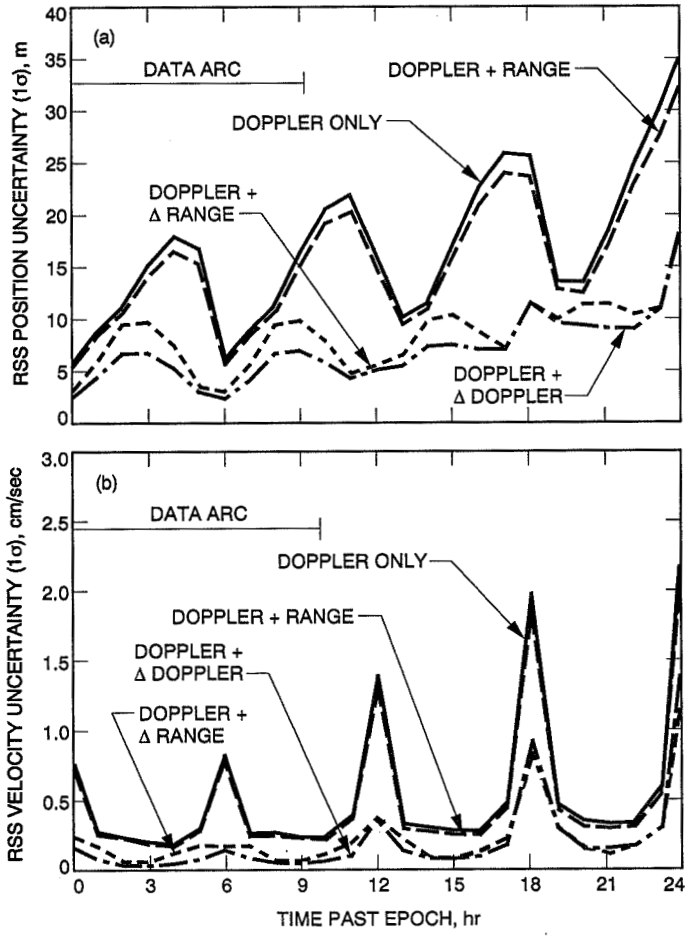


Fig. 5. VSOP 1-σ orbit determination accuracy statistics for: (a) expected RSS total position uncertainty; (b) expected RSS total velocity uncertainty.

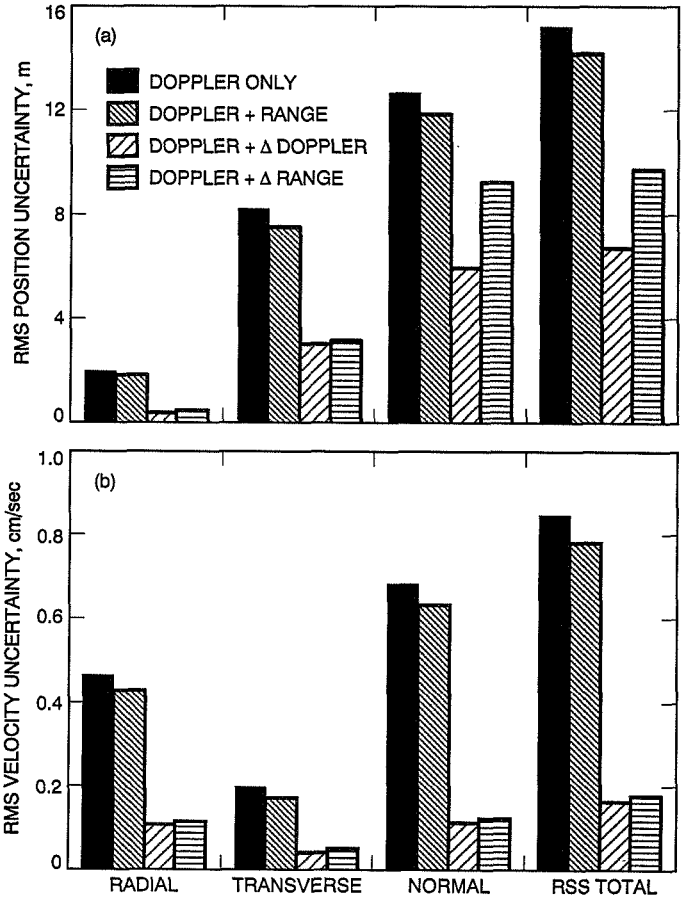


Fig. 6. VSOP 1-σ orbit determination accuracy statistics for: (a) expected position uncertainty at initial apogee crossing in terms of orthogonal components; (b) expected velocity uncertainty at first perigee passage (6 hr past epoch) in terms of orthogonal components.

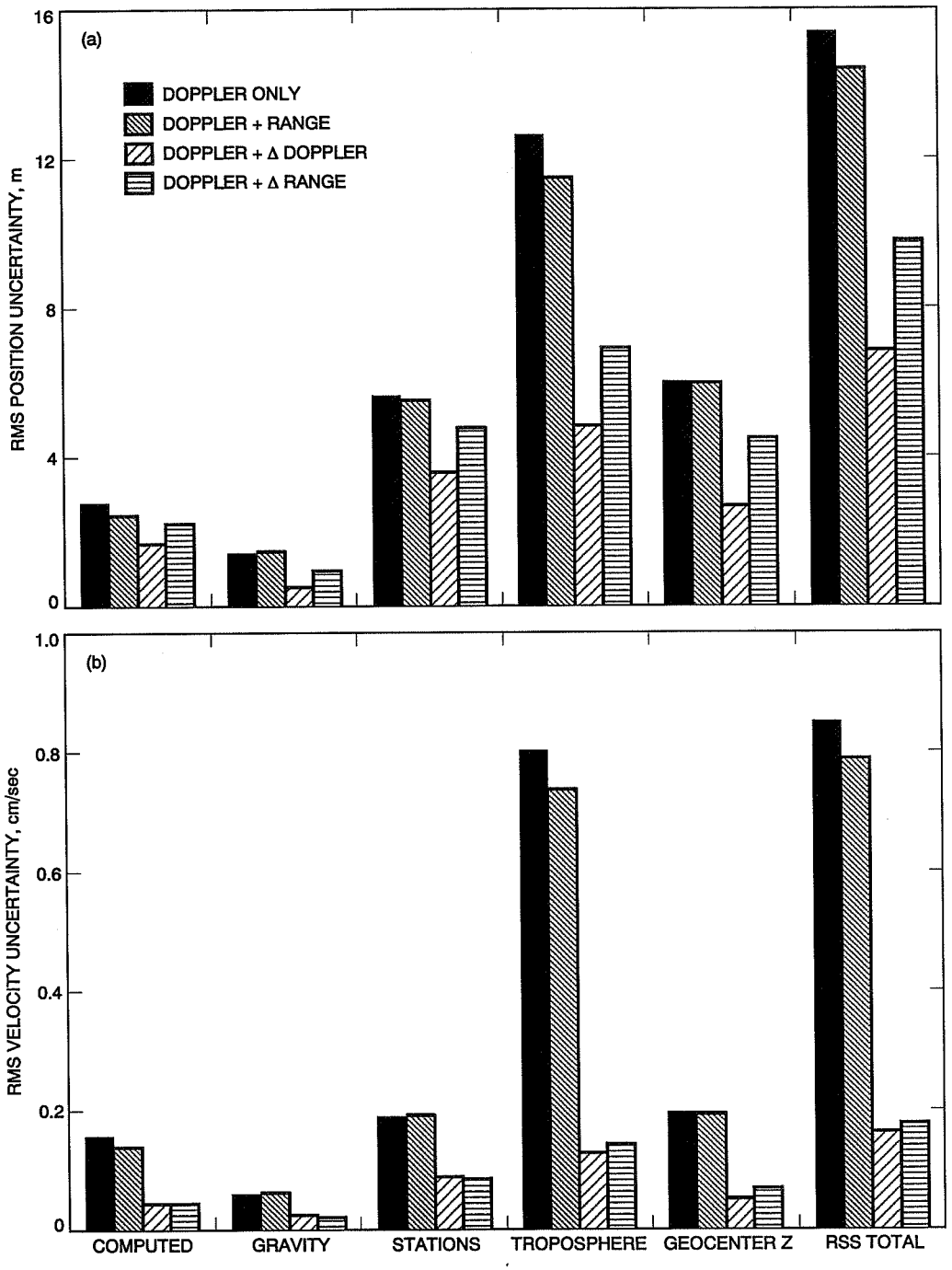


Fig. 7. VSOP 1-σ orbit determination accuracy statistics for: (a) expected position uncertainty at initial apogee crossing in terms of individual error sources: (b) expected velocity uncertainty at first perigee passage (6 hr past epoch) in terms of individual error sources.

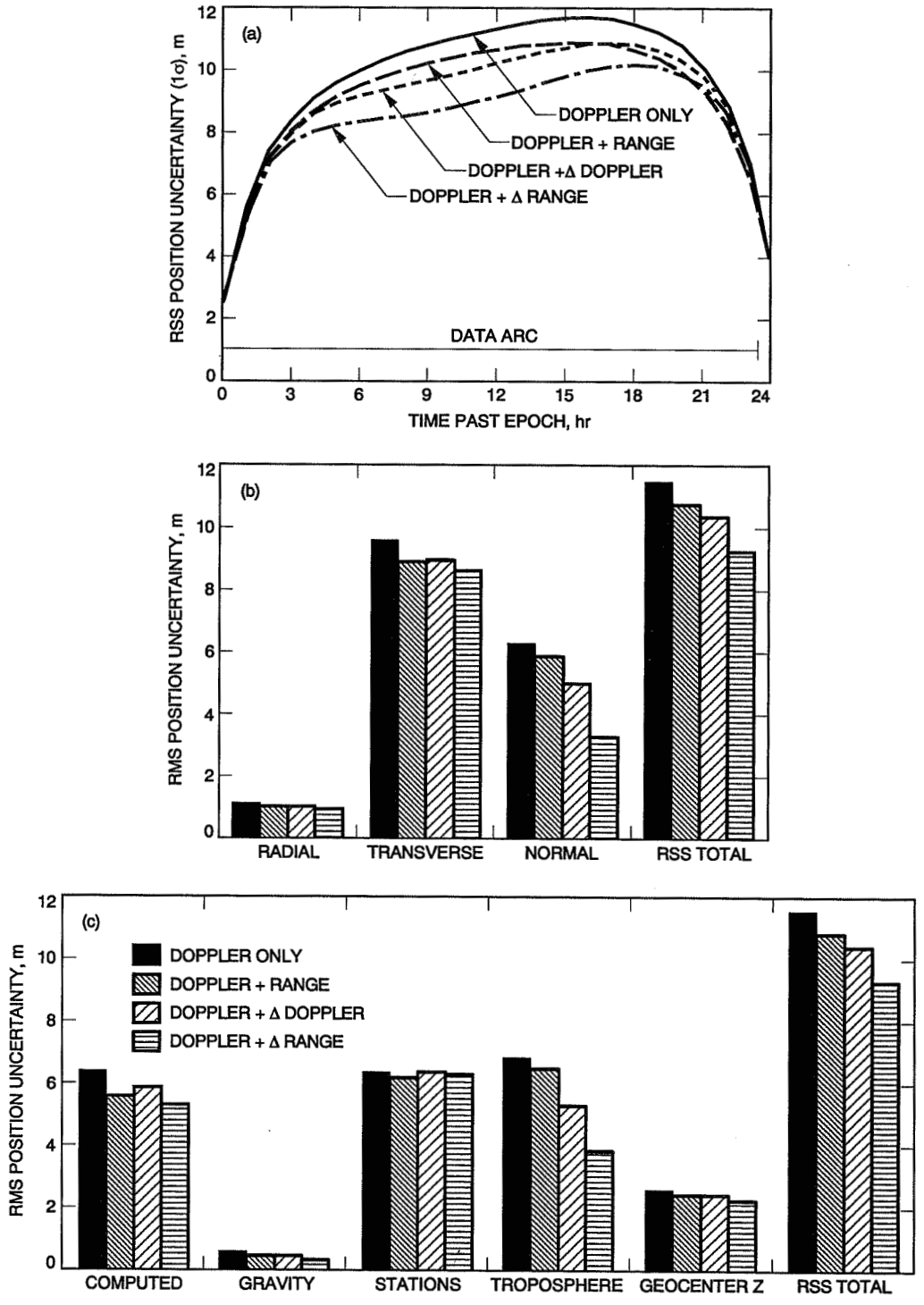


Fig. 8. RADIOASTRON 1-σ orbit determination accuracy statistics for: (a) expected RSS total position uncertainty; (b) expected position uncertainty at initial apogee crossing in terms of orthogonal components; (c) expected position uncertainty at initial apogee crossing in terms of individual error sources.

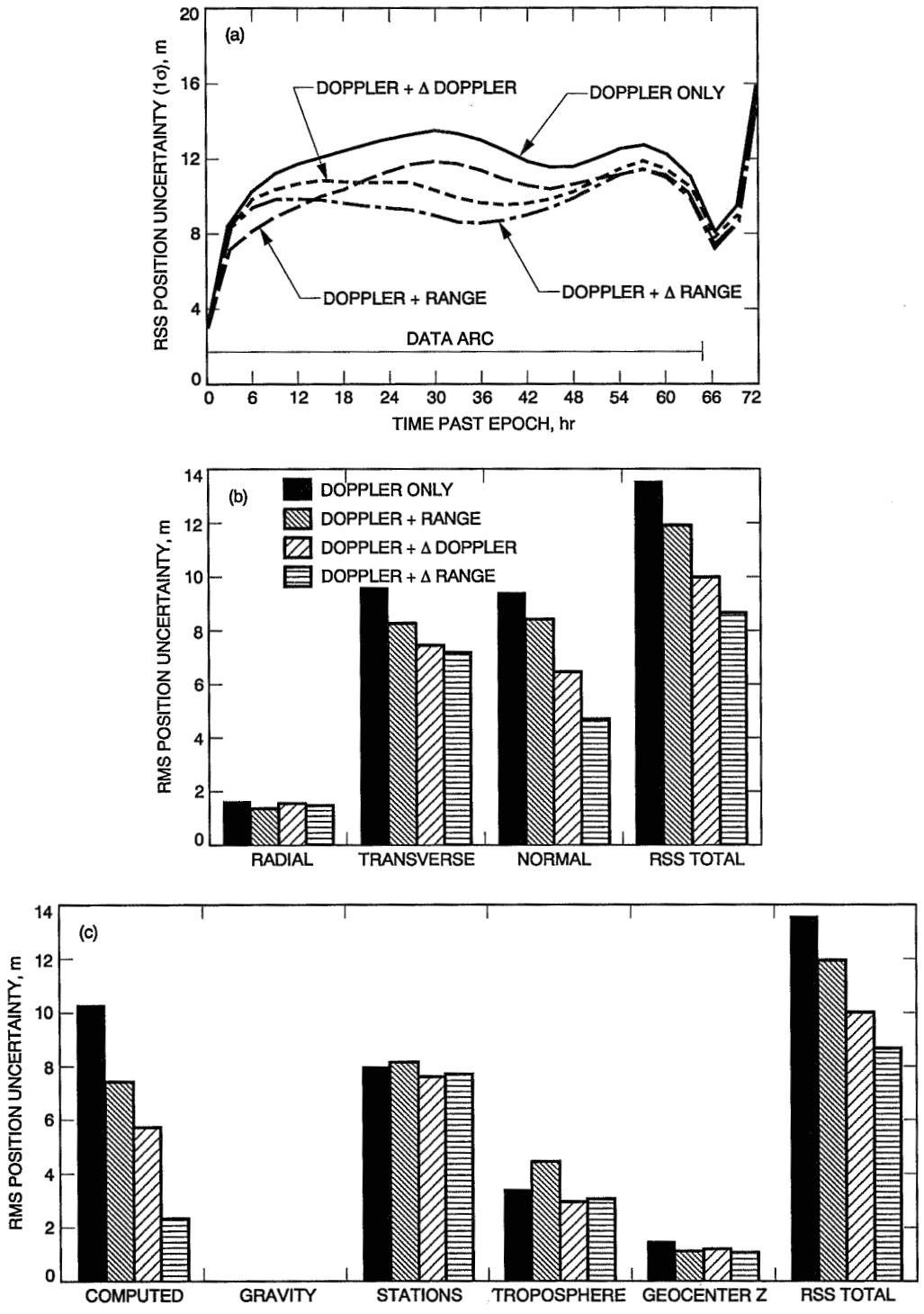


Fig. 9. IVS 1-σ orbit determination accuracy statistics for: (a) expected RSS total position uncertainty; (b) expected position uncertainty near initial apogee crossing in terms of orthogonal components; (c) expected position uncertainty near initial apogee crossing in terms of individual error sources.

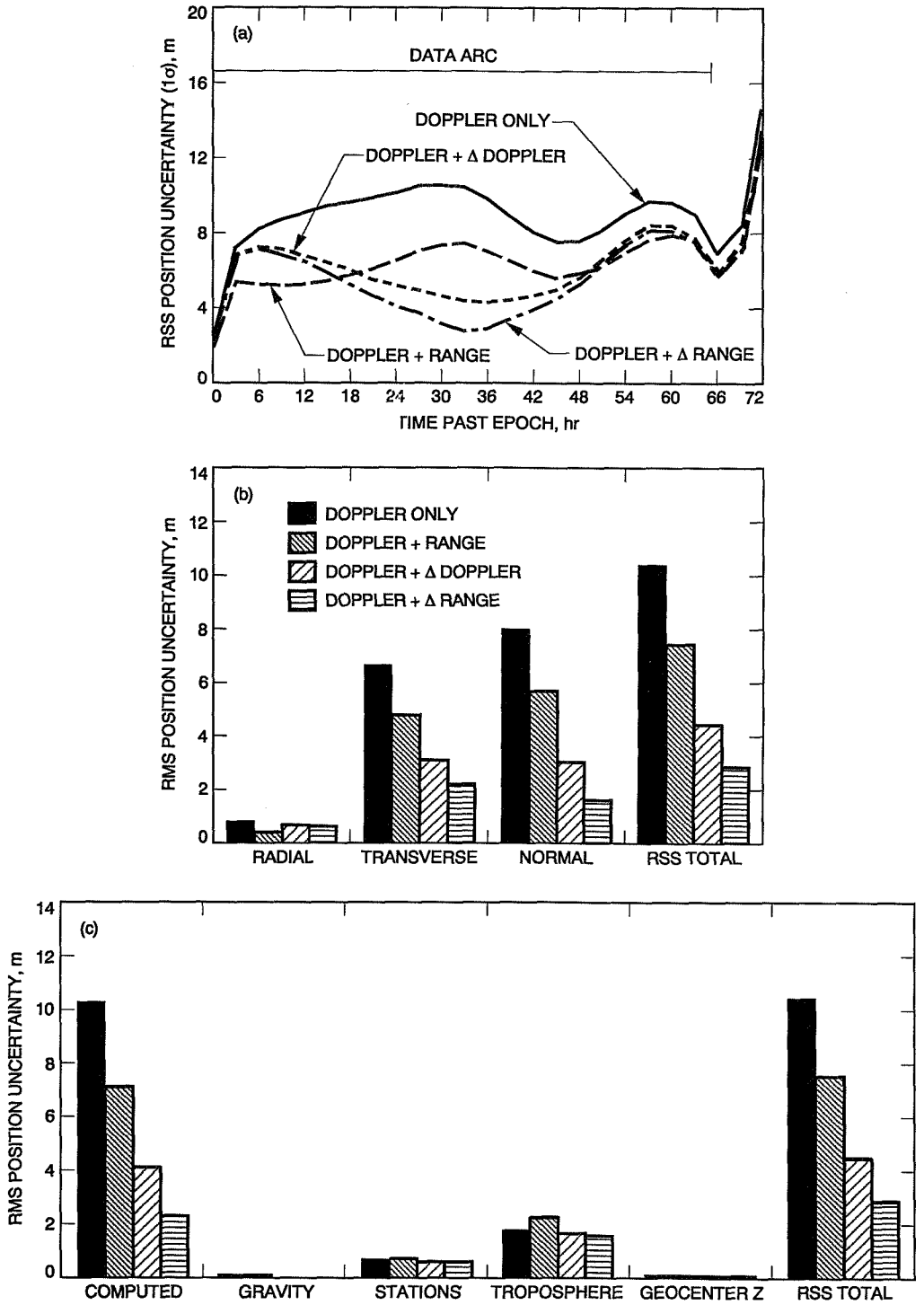


Fig. 10. IVS 1-σ orbit determination accuracy statistics using GPS-based ground observations for: (a) expected RSS total position uncertainty; (b) expected position uncertainty near initial apogee crossing in terms of orthogonal components; (c) expected position uncertainty near initial apogee crossing in terms of individual error sources.

Algorithms for Structural Natural-Frequency Design

R. Levy
Ground Antennas and Facilities Engineering Section

The algorithms used in the JPL-IDEAS antenna-structure-design-optimization program are furnished here. The algorithms are based upon the operational research method of optimality criteria and the structural analysis method of virtual work. Examples of the natural-frequency-constrained design of an antenna tripod structure are included.

I. Introduction

The JPL-IDEAS program [1] is a finite-element structure-design-optimization program with minimum structure weight as the design objective. In addition to the more conventional constraints, such as those on stresses and displacements, structures can be subject to constraints on antenna microwave performance parameters [2]. The program can also accept minimum structural natural frequency [3] for any specified vibration mode as a constraint. The design variables are the areas of rod members or the thicknesses of shear, triangular, or quadrilateral membrane plates. The design approach employs the Optimality Criteria Method [4] in which Lagrangian multipliers [5,6] are used to determine the sizes of the design variables, and a virtual work formulation is used to determine the sensitivities of the design variables.

Details of the natural-frequency design algorithm used in an earlier version of the JPL-IDEAS program are described in [3]. As explained there, depending upon details of formulation of the optimization problem, it was necessary to use the artifice of scaling the final design to obtain the minimum weight design that met the natural-frequency constraints. In other research that also concentrates on the natural-frequency design case [7,8,9], it is also necessary to resort to scaling or recursion to obtain the final design. More recently, the JPL-IDEAS natural-frequency design algorithm has been improved to eliminate the need for scaling the design. Scaling is especially objectionable if members are to be selected from tables of commercially available structural shapes. The current algorithm provides an explicit equation for computation of the Lagrangian multiplier in the case of a single-mode natural-frequency constraint, and it can be extended to deal with constraints on multiple-mode frequencies. This algorithm is described here.

II. Sensitivity Coefficient

To arrive at the sensitivity of natural frequency to the design variables, the natural-frequency eigenvalue of a particular vibration mode is expressed in terms of the Rayleigh quotient as

$$\omega^2 = \frac{\phi^t \mathbf{K} \phi}{\phi^t \mathbf{M} \phi} \tag{1}$$

in which ϕ is the mode-shape eigenvector, $\mathbf{K}$ is the assembled structure stiffness matrix, $\mathbf{M}$ is the assembled

structure lumped (diagonal) mass matrix, ω is the natural circular frequency, and ω^2 is the eigenvalue. Taking the partial derivative of the eigenvalue with respect to a particular design variable a_i (such as a rod area or plate thickness, either as an individual member or linked member group) and also using the customary definitions of generalized modal mass $\mathcal{M}$ and generalized modal stiffness $\mathcal{K}$, e.g.,

$$\mathcal{M} = \phi^t \mathbf{M} \phi \tag{2}$$

$$\mathcal{K} = \phi^t \mathbf{K} \phi \tag{3}$$

the partial derivative can be expressed as

$$\frac{\partial \omega^2}{\partial a_i} = \frac{1}{\mathcal{M}} \left[\phi^t \frac{\partial \mathbf{K}}{\partial a_i} \phi - \phi^t \frac{\partial \mathbf{M}}{\partial a_i} \phi \omega^2 \right] \tag{4}$$

in which the partial derivatives of the assembled stiffness and mass matrices can be expressed in terms of element member stiffness k_i and mass m_i matrices for the ith member as

$$\frac{\partial \omega^2}{\partial a_i} = \frac{1}{\mathcal{M} a_i} \left[\phi^t k_i \phi - \phi^t m_i \phi \omega^2 \right] \tag{5}$$

In the right side of the above equation, the first term in the brackets represents the virtual work and the second term the virtual kinetic energy for this element in this vibration mode. Note that each of the bracketed terms involves only the small subset of the eigenvector associated with the connectivity of the element. The bracketed term is replaced by the symbol V_i, which will be considered as an expression of vibratory virtual work, as modified by a subtractive kinetic energy term, e.g.,

$$V_i = \left[\phi^t k_i \phi - \phi^t m_i \phi \omega^2 \right] \tag{6}$$

Consequently, the partial derivative of the eigenvalue with respect to the ith member design variable is given by

$$\frac{\partial \omega^2}{\partial a_i} = \frac{V_i}{\mathcal{M} a_i} \tag{7}$$

It is convenient to design for structural natural frequency as much as possible within the code that is already in place for static loading design. The JPL-IDEAS program uses the virtual work of each design variable to design for static loading. In contrast to Eq. (6), this virtual work is expressed by a displacement-method formulation for plate members and by a force-method formulation for bar members as follows:

$$V_i = \phi_R^t k_i \phi_D = \left(C_R C_D \frac{\ell}{aE} \right)_i \tag{8}$$

in which the subscript R denotes a real external loading and D denotes a virtual external loading, ϕ_R represents the corresponding real displacement vector, and ϕ_D represents the virtual loading displacement vector. For a bar element, C_R and C_D represent real loading and virtual loading stress resultants, while ℓ, a, and E represent the corresponding length, area, and modulus of elasticity. To simplify the exposition, the remaining discussion will emphasize the treatment for bar members. Minor modifications appropriate to treat plate members can be found in [3].

In the case of bar members, it can be seen that the virtual work for the natural-frequency design can have the identical form of that in Eq. (8) for the static-loading design when using the second equality of this equation, provided that C_R is interpreted as the stress resultant corresponding to the displacements of the eigenvector, and C_D (dropping the subscript i) is computed as

$$C_D = C_R - \omega^2 \phi^t m \phi \frac{aE}{\ell C_R} \tag{9}$$

Furthermore, if, as in [6], a combined stress coefficient term FIJ' is defined as

$$FIJ' = C_R \frac{C_D}{E} \tag{10}$$

the virtual work for the ith element (again dropping the subscript) can be expressed as

$$V_i = \left(\frac{FIJ'\ell}{a} \right)_i \tag{11}$$

Consequently, the sensitivity of the eigenvalue to the ith bar design variable is given by

$$\frac{\partial \omega^2}{\partial a_i} = \left(\frac{FIJ'\ell}{a^2} \right)_i \cdot \frac{1}{\mathcal{M}} \tag{12}$$

At this point, it is convenient to replace FIJ' with

$$FIJ = \frac{FIJ'}{\mathcal{M}} \tag{13}$$

so that the sensitivity equation becomes

$$\frac{\partial \omega^2}{\partial a_i} = \left(\frac{FIJ\ell}{a^2}\right)_i \tag{14}$$

III. Optimality Criterion

Let ω^* be the minimum requirement for the natural frequency; the following constraint inequality then applies:

$$\omega^{*2} - \omega^2 \leq 0 \tag{15}$$

The object is to minimize the structural weight, which, for bar members with a weight-density parameter γ, is given by

$$\text{objective} = \min\left(\sum(\gamma \ell a)_i\right) \tag{16}$$

Forming the Lagrangian $\mathbf{L}^*$ in the conventional way, with λ as the yet-to-be-determined Lagrangian multiplier, provides

$$\mathbf{L}^* = \sum(\gamma \ell a)_i + \lambda(\omega^{*2} - \omega^2) \tag{17}$$

Setting the partial derivative of the expression with respect to each design variable equal to zero and using Eq. (14) provides the explicit expression for the optimum value of the design variable $\bar{a}_i$ as follows:

$$\bar{a}_i = \left[\lambda\left(\frac{FIJ}{\gamma}\right)_i\right]^{1/2} \tag{18}$$

The above equation defines the optimality criterion for each of the design variables. The remaining requirement is to determine λ.

IV. Lagrangian Multiplier

Using the sensitivity expression and summing over all the members, $\Delta\omega^2$ (the change in eigenvalue) can be estimated as follows:

$$\Delta\omega^2 = \sum\left[\left(\frac{FIJ\ell}{a^2}\right)(\bar{a} - a)\right]_i \tag{19}$$

Writing the right side of the above equation as the difference of two summations, the following is obtained:

$$\Delta\omega^2 = \sum\left(\frac{FIJ\ell\bar{a}}{a^2}\right)_i - \sum\left(\frac{FIJ\ell}{a}\right)_i \tag{20}$$

By reexamining Eqs. (6), (8), and (10) and with some algebra, it can be shown that the second summation in Eq. (20) can be written as

$$\sum\left(\frac{FIJ\ell}{a}\right)_i = \omega^2 \frac{\mathbf{M_F}}{\mathcal{M}} \tag{21}$$

in which $\mathbf{M_F}$ is the portion of the generalized mass matrix contributed by the nonstructural (parasitic) masses. That is, the generalized mass can be considered as the sum of the contributions from the structural-design variables $\mathbf{M_S}$ and the contribution of the fixed masses, $\mathbf{M_F}$; e.g.,

$$\mathcal{M} = \mathbf{M_S} + \mathbf{M_F} \tag{22}$$

Solving the constraint equation as an equality provides

$$\Delta\omega^2 = \omega^{*2} - \omega^2 \tag{23}$$

The first summation on the right side of Eq. (20) can be written as the sum of two terms; one term depends upon free design variables that can be determined according to the optimality criteria; the second term has bounds a_b for the design variables, such as a side constraint or a move limit. Therefore, this term can be written as

$$\sum\left(\frac{FIJ\ell\bar{a}}{a^2}\right)_i = \sum_{free}\left(\frac{FIJ\ell}{a^2}\left(\frac{FIJ}{\gamma}\right)^{1/2}\right)_i \lambda^{1/2} + \sum_{bounded}\left(\frac{FIJ\ell}{a^2}a_b\right)_i \tag{24}$$

Finally, some algebra based upon using Eqs. (20), (21), (22), (23), and (24) provides an explicit expression to evaluate the Lagrangian multiplier as follows:

$$\lambda^{1/2} = \frac{\left(\omega^{*2} - \omega^2 \frac{\mathrm{M_s}}{\mathcal{M}} - \sum_{bounded} \frac{FIJ\ell}{a^2} a_b\right)}{\sum_{free} \frac{FIJ\ell}{a^2} \left(\frac{FIJ}{\gamma}\right)^{1/2}} \qquad (25)$$

Consequently, Eq. (18) in conjunction with Eq. (25) constitutes the solution to the natural-frequency design problem for a particular natural mode. The foregoing developments could also be extended to provide the solution for simultaneous constraints on the frequencies of several modes. These equations are applied iteratively in a sequence of design cycles with move limits. Move limits compensate for linearizations inherent to the foregoing formulations.

V. Design Examples

A hypothetical test problem is chosen to demonstrate design optimization. The structure considered is a modification of the actual DSN DSS 13 antenna-reflector tripod structure. Figure 1 shows the antenna and tripod assembly just after completion in the summer of 1990. Figure 2 shows the layout and geometry of the isolated tripod structure component, which is designed with constraints on minimum natural frequency. The analytical model contains about 350 individual structural trusswork rod and plate members, which are linked into 40 distinct design variable groups. There are about 200 nodes and 450 unrestrained degrees of freedom. In these examples, members were selected from a continuous spectrum of available sizes. In practice, there is a program option that causes members to be selected from discrete tables of commercially available structural shapes.

The frequency constraints for the first three natural modes were postulated to be 2.5, 5.2, and 5.8 Hz. The starting design consisted of about 4.8 kilopounds (kips) of structure weight plus 4.5 kips of parasitic, nonstructure weight. All three frequencies at the start were each about 20 percent less than the constraint.

Figure 3(a) shows the frequencies achieved for independent designs for each of these modes. The designs of the second and third modes approximately satisfied the first mode constraint. The design for the second mode did not satisfy the third mode constraint, nor did the third mode design satisfy the second mode constraint. The first mode design did not satisfy either of the other two constraints. The structure weights achieved during these designs are shown in Fig. 3(b). It can be seen that both the second and third mode designs entailed substantial weight increases. Nevertheless, the first mode design resulted in a weight reduction while providing the desired increase in natural frequency.

In another example, all three constraints were applied simultaneously by using an envelope method as an approximation to a multiple-constraint design. This method treats the constraints sequentially and maintains the values of the design variables at a level no less than the values determined for previously treated constraints. Although not strictly an optimal procedure, it has often been found to work well in practice, especially for stress and displacement constraints. The history of the simultaneous designs and the structure weight is shown in Fig. 4. All the constraints are essentially satisfied at the ninth design cycle, and the structure weight is less than 3 percent greater than that of the isolated design for the third mode.

VI. Conclusion

The algorithm used to design for minimum natural frequency and in the JPL-IDEAS structural optimization computer program has been described and demonstrated by an example. The optimality criteria method, which is simple in concept and in execution, is employed. This method provides an explicit algorithm—almost trivial to invoke—to size the design variables. Determination of the Lagrangian multiplier, which is used in the algorithm, requires most of the computational effort. The algorithm is based on the well-known virtual work, dummy-load concept. The final design to meet the constraints is achieved directly and avoids the artifice of uniform scaling of computer-derived results.

Acknowledgment

The careful review and constructive comments of Michael Thorburn are appreciated.

References

[1] R. Levy and D. Strain, *JPL-IDEAS, Iterative Design of Antenna Structures,* COSMIC, NPO 17783, University of Georgia, Athens, Georgia, October 1988.

[2] R. Levy, "Optimization of Antenna Structure Design," *Proceedings of the Eighth Conference on Electronic Computation*, ASCE, Houston, Texas, pp. 114–129, February 21, 1983.

[3] R. Levy and K. Chai, "Implementation of Natural Frequency Analysis and Optimality Criterion Design," *Computers and Structures*, vol. 10, pp. 277–282, 1979.

[4] L. Berke and N. S. Khot, "Use of Optimality Criteria Methods for Large-Scale Systems," *Structural Optimization*, AGARD-LS-70, pp. 1-3–1-19, 1974.

[5] L. A. Schmit and C. Fleury, "Structural Synthesis by Combining Approximation Concepts and Dual Methods," *AIAA Journal*, vol. 18, no. 10, pp. 1251–1260, October 1980.

[6] R. Levy and W. Parzynski, "Optimality Criteria Solution Strategies in Multiple Constraint Design Optimization," *AIAA Journal*, vol. 20, no. 5, pp. 708–715, May 1982.

[7] V. B. Venkayya, N. S. Khot, V. A. Tischler, and R. F. Taylor, "Design of Optimum Structures for Dynamic Loads," *Proceedings of the Third Air Force Conference on Matrix Methods in Structural Mechanics*, Air Force Dynamics Laboratory, Wright-Patterson Air Force Base, Canton, Ohio, pp. 619–658, October 1971.

[8] R. Grandhi and V. Venkayya, "Structural Optimization with Frequency Constraints," *AIAA/ASME/ASCE/AHS, 28th SDM Conference,* Paper 87-0787-CP, Monterey, California, pp. 322–333, April 6–8, 1987.

[9] R. A. Canfield, V. B. Venkayya, and R. V. Grandhi, "Structural Optimization with Stiffness and Frequency Constraints," *Mechanics of Structures and Machines*, vol. 17, no. 1, pp. 95–110, 1989.

ORIGINAL PAGE
BLACK AND WHITE PHOTOGRAPH

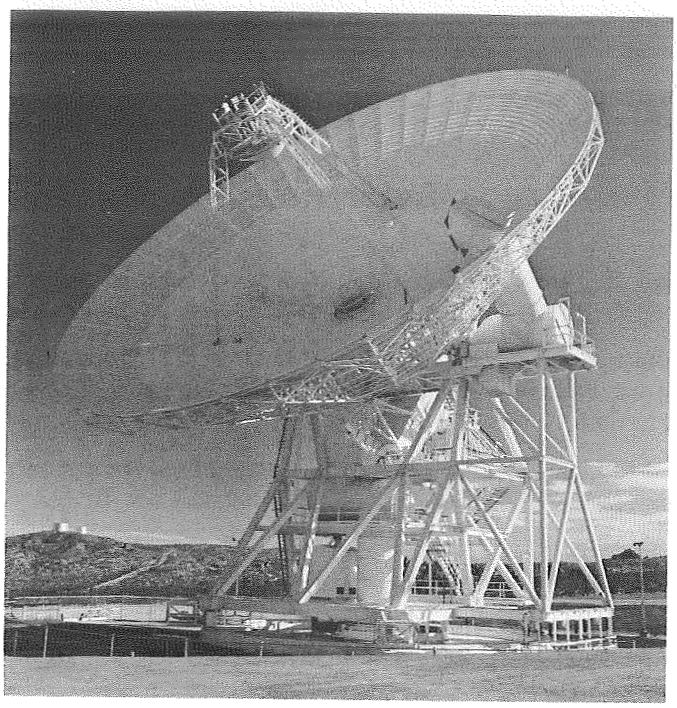


Fig. 1. DSN DSS-13 antenna-reflector-tripod structure.

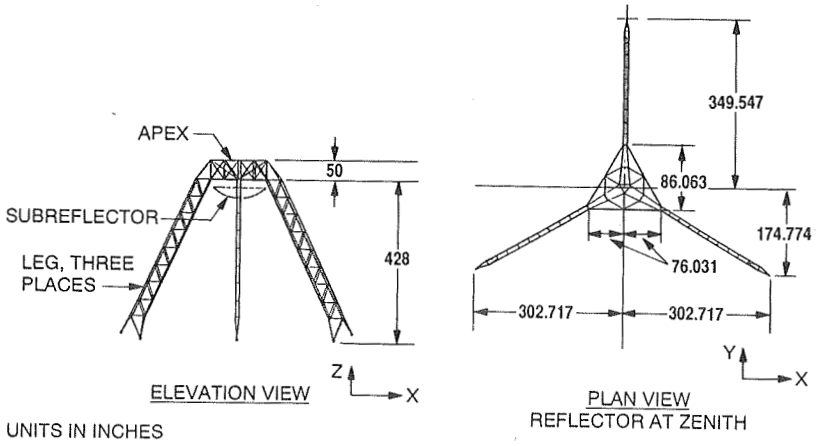


Fig. 2. DSS-13 tripod geometry.

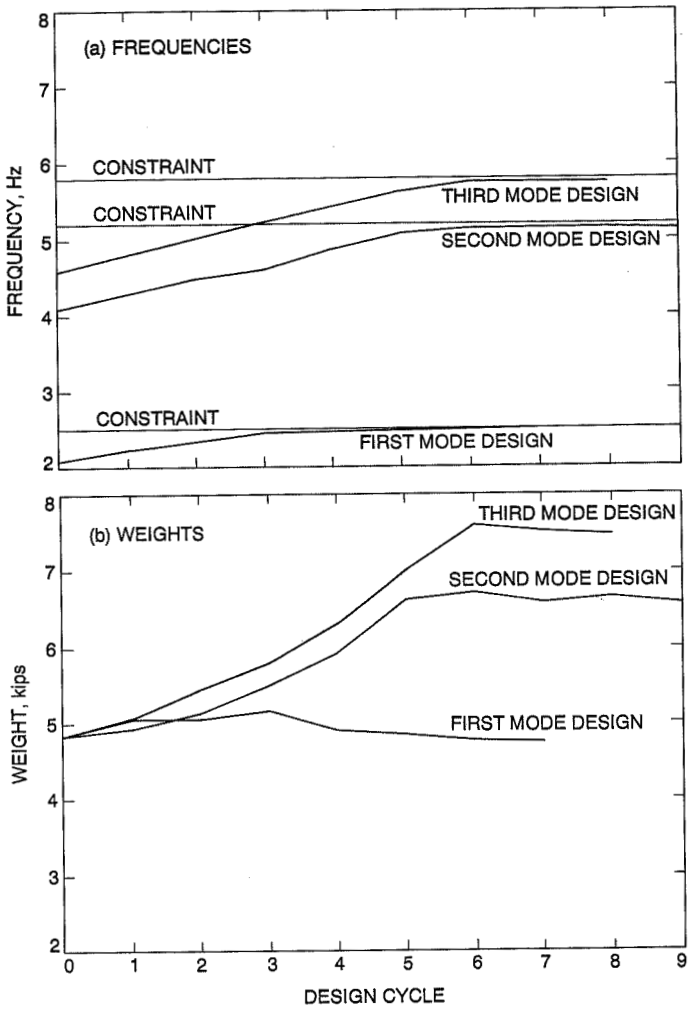


Fig. 3. Independent designs for single modes.

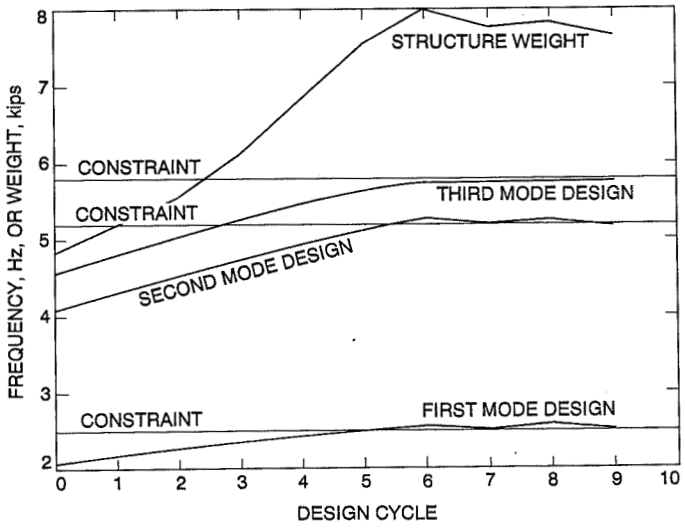


Fig. 4. Simultaneous constraints for three modes.

Pattern-Recognition Techniques Applied to Performance Monitoring of the DSS 13 34-Meter Antenna Control Assembly

J. Mellstrom
Ground Antennas and Facilities Engineering Section

P. Smyth
Communications Systems Research Section

This article describes the results of applying pattern-recognition techniques to diagnose fault conditions in the pointing system of one of the Deep Space Network's large antennas, the DSS 13 34-m structure. A previous article described an experiment whereby a neural network technique was used to identify fault classes by using data obtained from a simulation model of the DSN 70-m antenna system. This article describes the extension of these classification techniques to the analysis of real data from the field. The general architecture and philosophy of an autonomous monitoring paradigm is described and classification results are discussed and analyzed in this context. Key features of this approach include a probabilistic time-varying context model, the effective integration of signal processing and system identification techniques with pattern-recognition algorithms, and the ability to calibrate the system given limited amounts of training data. The article reports recognition accuracies in the 97–98-percent range for the particular fault classes included in the experiments.

I. Introduction

This article is intended as an overview of recent research on how to design a health-monitoring system for the DSN antennas. In particular, a systems perspective is provided on how a pattern-recognition component may be embedded within a more standard monitoring architecture. Not included are particular aspects of related technical topics such as classifier design, detection of change in a time series, etc.

The article develops as follows: First, the background to the problem is outlined. This is followed by a general discussion of fault diagnosis, with a description of what has become an evolving design for the autonomous monitoring

system. The conditions for the experiment are described, as well as how the data were collected and why the faults were chosen. The final section is a discussion of the experimental results and the various algorithms that were run.

II. Background and Motivation

The facilities of the DSN, such as the large antenna structures at the Goldstone, Madrid, and Canberra sites, represent a significant capital investment by NASA and JPL. These antenna structures are critical to the performance of the DSN, as they are potential single points of failure in end-to-end network operation. Traditionally, relatively low rates of data loss over the years have occurred due to the efforts of JPL engineering and operations personnel to ensure that problems have been identified and corrected quickly. As long as critical missions such as planetary encounters were of short duration and relatively sparse on the calendar, such labor-intensive approaches were feasible. However, the nature of JPL planetary missions is already changing to involve longer duration planetary encounters such as the Magellan project. In this context, round-the-clock manual supervision of the various antenna structures will not be feasible or economical, hence, the risk of science data loss will be significantly increased. Notwithstanding this change in the DSN operational environment, two other factors independently combine to strongly support the argument in favor of autonomous monitoring of DSN antennas. The first factor is the age of the 70-m antenna systems. The initial projected operational lifetime of these antennas has long been passed, and one can reasonably expect that system components will become more failure-prone as time progresses. The second factor is the planned shift in operational frequencies from X-band (8.45 GHz) to Ka-band (32 GHz); from an antenna-pointing standpoint, this necessitates much more precise pointing accuracies than are currently feasible. In this context, an antenna-monitoring system can, in principle, provide much more information than simply setting off alarms when faults occur. In particular, the monitoring system can complement existing controllers by providing on-line information about pointing-system components.

The above arguments in favor of autonomous health monitoring may be convincing in themselves. However, it should further be noted that recent advances in both hardware and algorithms are what really make an autonomous system feasible. Until recently, the computational capabilities required for real-time data acquisition, signal processing, and pattern recognition for this type of problem would have required large computers at significant cost. A major goal is to produce a system that can be implemented at low cost (perhaps on a single board) using conventional software such as C. This makes for higher reliability and lower cost maintenance from an implementation standpoint. In addition to the recent hardware advances, there has been a resurgence of interest in pattern-recognition techniques and applications for autonomous systems over the last few years, driven largely by the promise of neural network approaches to the problem. From a development point of view this is advantageous, as the significant amount of theoretical and applications-oriented work being carried out in other research institutions can be leveraged.

III. Problem Description: Detecting Failures in Antenna-Control Assemblies

A. Are Antenna Faults a Problem in the DSN?

The importance of the antennas in terms of DSN operation was decribed in the introduction. One question that must be asked, however, is whether antenna subsystems are prone to failure. Direct evidence is difficult to acquire since there is no DSN database that tracks equipment outages at the component level. The easiest information to obtain is that based on discrepancy reports. There are limitations on what can be inferred from these data. In particular, the data outages reported are only a lower bound on the length of time a particular subsystem (and thus a whole antenna) is not available to support tracking. Other inaccuracies can be caused by the reporting mechanism where reports written against the antenna system are due to problems such as wind and snow or procedural errors, which are of no relevance in determining antenna system hardware reliability, although they are relevant for availability.

Nevertheless, one can look at the *relative* magnitudes of data outages caused by the different subsystems in a DSN ground station. Over the period January 1986 to March 1991, 21.9 percent of the total data outage hours (as measured during scheduled tracks) for the DSN as a whole were written against the antenna mechanical subsystem (ANT). The antenna system was the second most likely culprit in this regard after the DSCC Telemetry Subsystem (DTM), which accounted for 28.2 percent of lost data hours, with radio frequency interference (RFI) in third place at 16.3 percent. The top five subsystems account for over 87.7 percent of the total data hours lost. Data outage information is summarized in Fig. 1. In absolute terms, 0.67 percent of all scheduled data hours over the past 5 years was lost because of problems written against the antenna mechanical subsystem. From these data, it seems reasonable to infer that antenna system problems

are a significant contributor to DSN downtime. More detailed information of this nature needs to be acquired, such as outage rates for specific antennas, component-level maintenance information, and so on.

B. Fault Detection and Diagnosis at Present in the DSN

Typically, a faulty component will manifest itself in the antenna-pointing system indirectly via a change in the characteristics of observed sensor points in the control loop. Because of the nonlinearity and feedback present, direct causal relationships between fault conditions and observed symptoms are difficult to establish. This can make manual fault diagnosis a time-consuming and inefficient process. In addition, if a pointing problem occurs while a spacecraft is being tracked, that antenna is often shut down and the track is handed over to another antenna, if possible. Hence, at present, diagnosis often occurs after the fact, where the original fault conditions may be difficult to replicate.

IV. Design of a Fault-Diagnosis System

A. Fault-Diagnosis Techniques

There are a variety of technical approaches for building a fault-diagnosis system. Three such approaches are considered here: control theory, artificial intelligence, and pattern recognition. Willsky [1] described some classic control-theoretic techniques for on-line diagnosis. In general, this approach requires a detailed observer model or filter for each type of postulated fault. This presents a problem, because constructing the observer models is quite difficult due to the high order and nonlinearity of the antenna drive system. In addition, the technique should be adaptive and able to recognize when new faults occur. These capabilities are difficult to incorporate into the conventional observer models. However, the problems are not insurmountable, and the control-theory techniques may offer the most straightforward solution for certain classes of easily identifiable faults.

Artificial-intelligence techniques for fault diagnosis can be classified into two categories. The first is rule-based expert systems. As pointed out in [2], systems with temporal behavior are poorly modeled by the rule-based paradigm. In addition, there is little in the way of experiential knowledge with which to build a rule base for the problem. The second category is that of model-based reasoning, where a high-level qualitative model is built to enable problems to be solved in a manner similar to reasoning from basic principles. While this approach holds promise, especially where novel faults may be expected to occur, the technology is not yet mature enough for application (see [2] for a more detailed discussion).

The third general approach is based on pattern recognition or machine-learning ideas. The idea is simple enough: Design an algorithm that can learn a model directly from observing system behavior. A learning approach precludes the necessity for considerable prior domain knowledge, which is a characteristic of both the other approaches mentioned above. Hence, the system can in some sense be "bootstrapped" into operation. It is a well-known fact in pattern recognition, however, that providing the appropriate domain knowledge to the learning algorithm (say, in the form of preprocessing the data) can considerably improve its performance. Initial investigations bear this point out, and it appears that a hybrid system, which uses components of all three approaches described here, is the most practical and promising avenue. Initial work has been with the pattern-recognition component of the model, which is the primary focus of the remainder of this article.

B. System Design Issues

The approach taken in system health monitoring can be considered to be a rough engineering analogy to passive human sensory and perception capabilities. In particular, there is a hierarchical pattern of information flow, as the raw time-series data are transformed and abstracted by various specialized modules, mapped into a categorical representation of variables of interest (such as fault classes), and finally integrated over time to enable reliable and robust decision making. This architecture is quite generic to systems that must passively sense and perceive their environment in real time, and is a natural choice based on purely engineering considerations.

Figure 2 shows a diagrammatic representation of information flow in the present system. The lower level modules perform dedicated signal-processing tasks to reduce the effective dimensionality of the incoming data. The next level of processing consists of pattern classification operating at a much slower rate in a lower dimensional space. In pattern-recognition terms, there is a feature-extraction stage followed by a probabilistic classifier operating in a continuous real-time mode. Experiments have striven to represent the class information by using probabilities rather than the less informative 1-of-N label information. In this sense, the output of the pattern-recognition component can be viewed as a *probabilistic state vector* that represents the system's best estimate of the state of the system at time t. The decision making component is isolated

to another level of processing after the pattern-recognition component. Note that unlike many applications where decision making and pattern recognition are combined via the use of loss functions, it better serves the purpose here to defer decision making. The separation of probability estimation and decision function allows each component to be modified independently. This is a very important practical design consideration, which allows the decision function to be changed without having to relearn the pattern-recognition component.

V. The DSS 13 34-m Beam-Waveguide Antenna Experimentation Testbed

A. Rationale for Experiment

Having reported initial successes on fault classification by using data from a simulation model [3], the natural next step was to see if similar success could be achieved by using actual data from the field. Obtaining real fault data presents something of a problem, since faults by their nature are unpredictable. Fortunately, with access to the new 34-m beam-waveguide (BWG) antenna at DSS 13, the somewhat bold step of creating faults in the antenna-pointing system in a controlled manner could be taken. One realizes immediately the limitations of this approach:

(1) The procedure cannot readily be repeated on operational antennas, in particular the 70-m antennas.

(2) Only known faults can be simulated, while in a real system, faults of the unknown variety are of great interest.

These limitations are discussed in greater detail later. At this point, it is sufficient to note that despite the limitations, the experiment represents a critical test in demonstrating that pattern-recognition technology can make a real contribution to solving the problems of fault diagnosis.

B. The System Under Study: The 34-m BWG Antenna Elevation Axis Drive Assembly

The antenna elevation axis drive assembly is a closed-loop control system that consists of two 7.5-HP DC motors, servo amplifiers, cycloid gear reducers, tachometers, and electronics for signal conditioning and servo compensation. A simplified block diagram of this system is shown in Fig. 3. A brief description of the system is provided here. Greater detail is provided in [4].

There are two inputs of interest in the antenna drive assembly: the rate command $\dot{\theta}$ and the bias command v_{bias}. In Fig. 3, the rate command $\dot{\theta}$ is applied by the antenna servo controller (ASC). (The ASC is the computer that controls the pointing of the antenna.) The rate command is filtered by the reconstruction filter G_c, then applied to the two subsystems labeled G_o. The term G_o represents two similar sets of amplifiers, motors, gear reducers, tachometers, and compensation electronics, and G_o is itself a closed-loop system referred to as the rate loop. Together, these two subsystems drive the antenna structure.

The remaining block in Fig. 3 is the torque share/bias regulator G_{ct}. This is a regulator circuit that has two functions. The first is to share the torque between the two motors and reduce the effect of parameter variations between them. The second function is to provide a torque bias between the two motors. The torque bias is very important in reducing the nonlinearities and improving gear reducer stiffness. The torque bias removes backlash and shifts the operating point of the cycloid gear reducers into a near-linear, high-stiffness region. The magnitude of the bias is controlled by the bias command.

C. Description of Measured Faults

Five faults were introduced into the antenna drive assembly, all in the drive electronics. The drive electronics were identified as the safest, most easily controlled, and least disruptive location to introduce faults into the drive assembly. Figure 4 shows a block diagram of one motor, amplifier, and gearbox set. Four of the faults were introduced into the rate-loop compensation and tachometer feedback, and the remaining fault was introduced into the torque share/bias feedback loop. During the measurement process, an additional fault appeared in the encoder. Since this was a real fault, it was included as an additional class. The six classes of failure are described below.

(1) The first fault was a noisy tachometer. A uniformly distributed, white-noise process was introduced into the tachometer feedback path of one rate loop, as shown in Fig. 4. This simulates the wear of the commutator or the tachometer bearings, and is a failure that commonly occurs in the DSN and degrades antenna pointing.

(2) The second fault was the total loss of one tachometer. A failure such as this is masked during normal operation because of the interaction of the two rate loops through the torque share/bias loop, and because the closed position loop remains stable.

(3) The third fault was the loss of the torque share/bias feedback to one of the motors. This was implemented by eliminating the signal labeled v_b in one of the summing junctions shown in Fig. 4.

(4) The fourth fault was the elimination of the integrator in the rate-loop compensation, block G_1 in Fig. 4. This effectively changed the rate loop from a Type 1 servo to a Type 0 servo.

(5) The fifth fault was the short circuit of the rate-loop compensation. Referring to Fig. 4, this was equivalent to setting $K_r = 1$, and $\tau_2 = \tau_3 = 0$.

(6) The sixth fault was a real encoder failure. The failure manifests itself as sudden jumps in the measured elevation angle. The magnitude of the jumps was 10 to 20 mdeg. It is hypothesized that the failure is due to binding in the encoder mechanism due to unforeseen radial loading.

Closer analysis of the data revealed that the encoder failure fault was *intermittent* in nature. Detection and classification of this fault is not difficult when using a standard statistical model based on monitoring encoder deviations from the mean expected value. However, this article focuses mainly on the pattern-recognition aspect of the problem. Hence, from this point onwards, only the *persistent* classes of faults are discussed.

D. Data Acquisition at DSS 13

Instrumentation was set up at DSS 13 to measure selected digital and analog signals from the elevation axis antenna drive assembly. These signals were the elevation encoder, current in each DC motor, bias command, rate command, the tachometers for each rate loop, and an average tachometer signal. Also monitored were two anemometers providing wind speed and direction information. Since wind is a nonstationary disturbance, it is important to include this measurement. This provides the environmental context within which nominal performance must be judged. For example, the antenna pointing performance that must be considered nominal subject to a gusty 30-mph wind could be the same as the performance during a calm day with degradation due to a drive fault. On all three days of data acquisition described below, the measured wind speed was less than 5 mph.

Data were collected on three days in January and February 1991. Nominal performance data were collected during antenna calibrations performed on January 25. The data were logged while the antenna was boresighting a radio source. This implies that while the antenna was tracking the radio source, step offsets in elevation and cross-elevation were periodically introduced.

Data were collected under nominal and induced fault conditions on February 11 and 22. The data of February 11 were collected during a simulated high-elevation track. A rate offset of -1 mdeg per sec was introduced. The faults were introduced independently. All the data were collected at elevation angles greater than 60 deg. At elevation angles less than 60 deg, the encoder failure (described above as fault 6) occurred. This precluded the possibility of making any low-elevation measurements.

The data of February 22 were collected while tracking a rising radio source. Again, the faults were introduced independently. Boresighting did not occur during the measurement period. All the data were collected at elevation angles less than 30 deg. Note that all data channels were sampled at 50 Hz to a resolution of 16 bits.

VI. Results of Pattern-Recognition Experiments

The focus of this article is primarily on the pattern-recognition component of the monitoring system design. As outlined earlier, this component is but part of an overall hierarchical strategy for autonomous monitoring.

A. Choice of Pattern-Recognition Technique

Statistical pattern-recognition techniques have been studied since the early 1960s when the availability of computer hardware made it possible to implement computationally demanding classifier design algorithms. A classifier can be thought of as being very similar to a regression equation model, except that instead of trying to predict a real-valued variable, one instead seeks to predict a *categorical* variable, where the categories are the class labels. Typically one has a feature space of K variables; the general idea is to label or partition the K-dimensional feature-space in such a way that an unlabeled feature vector or measurement can be assigned a class label. The standard procedure is to be given a *training set* of labeled sample data, namely N feature vectors or measurements which have class labels attached to them. From the labeled training data, one must *infer* a general relationship between the K features and the class labels: this relationship forms a decision rule or classifier. Let the class labels be c_i, $1 \leq i \leq m$, where m is the number of classes. If the classifier is to return a class label, c_i, without any indication as to confidence of the classifier in this decision, then this is a *discrimination* problem. Denote the K-fold feature vector by the vector random variable $\underline{X}$. If the classifier is to produce the output $p(c_i|\underline{x})$ for each class and any particular feature $\underline{x}$, then this is an *estimation* problem that involves multivariate estimation of the class-feature conditional probability density.

Classifier design techniques for the discrimination and estimation problems fall into two broad categories: parametric and nonparametric. The parametric approach usually involves an assumption that the class-conditional feature densities can be modeled as multivariate Gaussian. Under this assumption the optimal decision boundaries can be found as a function of the means and covariance matrix. The problem with this approach is that Gaussian distributions are often not a good model, and accurate estimation of the components of the covariance matrix requires a large amount of data. The nonparametric approach does not seek a direct parametric form for the conditional densities. Nearest-neighbor techniques, for example, seek to approximate the local value of the density function as a function of the "neighbors" of that point with a given class label. Parzen windows use the same notion of local estimates where the class-probability estimate is based on the weighted contribution of other data points of the same class label, by using various kernel functions. Decision-tree techniques seek the decision boundaries by partitioning the feature space using hyperplanes parallel to the feature axis in a hierarchical manner. All these techniques suffer some drawbacks, such as implementation complexity (nearest neighbor), poor scaling performance as a function of feature dimensionality (Parzen windows), and limited expressive capability (decision trees).

A recent technique that has attracted considerable interest for application to classification problems is that of feed-forward multilayer neural networks. Internal details of such models were outlined in a previous article [3] and some of the details of a particular three-layer network are reproduced in Appendix A. The network implements a set of nonlinear equations that models the relationship between the features and each of the output classes. The network weights play the role of coefficients in the nonlinear equations, and one can view the internal nodes of the network as implementing basis functions. Miller, Goodman, and Smyth [5] have shown how the objective function (used as an error metric to find the optimal set of weights) relates to maximum-likelihood and maximum a posteriori estimation. Similarly, Gish [6] provides a useful discussion about maximum-likelihood properties of network-training algorithms and relates network models to standard multivariate logistic regression models. Among the useful properties of networks are their universal approximation capabilities. Cybenko [7] and Hornik, Stinchcombe, and White [8] have shown that a network with a single hidden layer (a layer of nodes between the input and output) can approximate any continuous function to any desired degree of accuracy. This result is more of theoretical interest than of practical use since it does not prescribe how to find such a network; nonetheless, it demonstrates the important point that network models are a very powerful and flexible basis for approximation.

The problem of network design is then to find a suitable architecture (the number of layers in the model, the type of nonlinearity used) and a good set of weights for the network. The architecture selection problem is still largely done in a fairly ad hoc manner. Typically, a network with a single hidden layer is used with roughly twice as many hidden units (nodes in the hidden layer) as there are classes. Empirical results indicate that network performance is often relatively insensitive to the number of hidden units used (provided that one has a reasonable amount of training data and the number of units is not too small), which indicates a robustness relative to architecture. The number of input units and output units is set equal to the number of features and classes, respectively. Typically, the node "basis functions" are chosen as sigmoid functions (see Appendix A), but any smooth differentiable function can be used as far as the backpropagation training algorithm is concerned.

One of the key contributions to the resurgent interest in neural network models in recent years is the backpropagation algorithm [9], which provides an iterative method for finding a set of network weights that corresponds to a local maximum of the objective function. The algorithm is really nothing more than the application of the chain rule for derivatives applied to a graph structure, coupled with the notion of gradient descent in weight-space. While no tractable techniques are known to exist that guarantee that an optimal or near-optimal solution will be found, practical experience with the algorithm has by and large been remarkably successful. Indeed, while the field of neural networks in general involves a wide variety of biologically inspired computational models, most of the engineering applications of neural networks (and especially the successful ones) hinge on the ability of the backpropagation algorithm to find powerful nonlinear models from data [10–12]. For the experiments described here, an enhancement to the basic gradient descent technique for finding a good set of weights was used, namely, conjugate-gradient optimization, which accelerates the convergence to a solution. More details on this algorithm are presented in [3].

Neural network models offer a powerful new technique for finding classification and prediction models from data. Comparative studies in the literature have consistently shown that network models are as good as, or outperform, other standard pattern-recognition algorithms [13,14], so that network models are now often the model of choice for nonparametric pattern-recognition problems. The significant advantages of a network model are its low complexity

and speed if implemented in hardware, which is important in a real-time application such as this. It is worth noting that multilayer feed-forward network models for classification are finding their way into a myriad of applications in a manner *independent* of the original biological motivations for the model. In other words, as an engineering tool, network models have established themselves in their own right. The application of these models to antenna fault classification is decribed below.

B. Feature Selection and Generation

As described earlier, each data set consists of 12 channels of 16-bit data, sampled at 50 Hz for roughly 5 min for each fault. Hence, for each fault, in pattern-recognition terminology, there are 15,000 12-tuple feature vectors. While this might seem like a large data sample, in actual fact it is really only a very brief snapshot of antenna system data.

Using the raw time-series data directly as input to the classifier is obviously not the best approach to the problem. A hint can be taken from the way a human would discriminate among the classes by using gross structural features of the waveform to characterize it. With this in mind, it was decided that some simple useful features could be extracted directly from the time-series data. (Such "time-domain" features were succesfully used in discriminating fault data from a simulation model of a 70-m antenna [3]). Generation of these features first involved segmenting each channel of the time-series data into windows. The window size was chosen to be of 4-sec duration (200 samples) to give reasonably accurate estimates of the various features. The features consisted of order statistics (such as the range) and moments (such as the variance) of particular sensor channels. A set of seven features was selected which were judged likely to have good predictive power for the problem: motor current range and variance, countertorque range and variance, the range and variance of the average-tachometer sensor, and the variance of the difference between the two tachometer sensors. Even though some of these features are highly correlated, the redundancy is useful to combat any uncorrelated noise that may be present.

The plots in Figs. 5(a) and (b), 6(a) and (b), and 8 all follow the same convention, where the various parameters of interest are plotted versus a window index that corresponds to 4-sec time increments. One can imagine the horizontal axis to be a function of time. The correspondence among the faults and various regions of the axis works as follows: noisy tachometer (1–75), tachometer failure (76–150), bias loss (151–225), integrator short circuit (226–300), compensation short (301–375), and normal (376–450). The entire width of the horizontal axis corresponds to 30 min worth of data.

Figures 5(a) and (b) show plots of the discrimination capability of two of these features on data from "day2" (February 22): the variance of the average tachometer sensor and the motor current variance. Notice that each feature has the ability to discriminate some of the classes and not the others. For example, motor current variance can discriminate the tachometer faults in windows 1 to 150 from the other faults but not from each other, and it can also discriminate the compensation short fault. Notice also that in the variance of the average tachometer sensor there are noise spikes in the data. The motivation for not seeking a completely minimal set of discriminatory features is to retain some redundancy and, hence, robustness in the presence of such noise.

An autoregressive (AR) modeling technique was applied to the motor current sensor. In particular, a variation of autoregression referred to as ARX was used [15], where the "X" refers to an exogenous input variable to the system. The system is modeled by using

$$y(t) + \sum_{i=1}^{p} a_i y(t-i) = \sum_{j=1}^{q} b_j u(t-j) + e(t),$$

$$t = 1, 2, \ldots, N$$

where $y(t)$ is the motor current, $u(t)$ is the system input (in this case, the rate command sensor), $e(t)$ is an additive white-noise process, and a_i and b_i are the model coefficients. A model with $p = 5$ and $q = 3$ was chosen since it provided the best trade-off between goodness-of-fit and model complexity. Note that the more traditional AR model assumes that the system is driven by a white-noise process, i.e., the input $u(t)$ would be replaced by a noise term in an AR model. A 4-sec window was used to segment the data and the ARX coefficients were estimated on a block-by-block basis. For on-line implementation, a standard *recursive* estimation scheme can be used. Figures 6(a) and (b) show the discriminatory capability of the first and second ARX coefficients as plotted for day2 data. Here it is seen that the first coefficient alone is almost sufficient to linearly discriminate among the classes; however, there is some overlap between the tachometer faults, and again between tachometer failure and normal conditions. It is not surprising that the ARX coefficients are useful discriminants since under appropriate conditions they are a sufficient statistic for the data, i.e., they represent all

the information in the data. Consequently, one can expect them to provide more useful information than the simpler time-domain features.

Analysis of the data revealed that the bias loss and integrator short-circuit faults were completely indistinguishable from the normal case. It was expected that some variation would be discernible given these faults; however, it appears that the system is robust enough to withstand such failures. Hence, the data for these two faults were relabeled as normal.

C. Multiple-Network Architecture

Experiments were carried out using a hierarchical set of networks. Specifically, two networks were trained: one on the time domain, the other on the ARX model features. Then the outputs of these two networks were fed into a third "judge" network, which is again trained using the same class labels as the original data. Figure 7 shows the overall model. The intuitive notion behind this multiple-network architecture is that the "judge" network can learn to discriminate which of the first two specialist networks are reliable on which classes. In particular, it learns how to weigh the estimates provided by the first two networks. A useful analogy might be a manager who has some technical experts on staff and needs to weigh their opinions on particular issues, taking care to note the strengths and weaknesses of each. This approach can be seen as a version of the mixture models proposed by Jacobs, Jordan, Nowlan, and Hinton [16]. One can expect this task decomposition approach to work in situations where little gain can be expected from directly combining the features in the two sets. In another sense this can be viewed as one large network where subsets of the feature space are not combined in the early layers to reduce the effect of wasting network resources. This is a more effective and efficient method than simply combining all the features into a single, large, fully connected (between the layers) network, as has been found in numerous large-scale network applications [17–20].

D. Classification Results

As described earlier, the data were collected at DSS 13 on January 25, February 11, and February 22, 1991. This section focuses on the data sets obtained on February 11 and 22, referred to as "day1" and "day2," respectively (no faults were induced on January 25). The goal of the classification experiment was to see if data from each day could be used to predict conditions on the other day. Hence, the methodology can be considered a simple two-way validation test. The results are shown in detail in Table 1. The results for testing on day1 and day2 imply that the models were trained on data from day2 and day1, respectively. Each component network was a three-layer model. The number of hidden units was fixed at 8 for each of the time-domain and ARX networks, at 10 for the network using both sets of features, and the "judge" component of the multiple-network architecture had 8 hidden units.

On average, the multiple network did better than the other models, but the difference is slight. Clearly, most of the useful discrimination ability is in the ARX coefficients rather than the time-domain features, there being an 8-percent difference in mean classification accuracy between these two schemes. These results are quite good if one takes into account that no time correlation is used, i.e., each classification decision is made independently of the other. Clearly, it is preferable to have the scheme correlate its decisions in some manner. In effect, one wishes to model the prior belief that faults are persistent over time and are not likely to change from one 4-sec window to the next. A scheme to do this time-dependent decision making is described in the next section. Note that a popular approach in sequential pattern-recognition problems of this nature is to try and estimate the time dependency from the data. This is usually achieved by providing as inputs to the model (a network in this case) not only feature and class labels at the present time, but also the labels and possibly the features from some window into the past. Such a scheme is viewed as unnecessary in this application since it would make the model much more complicated. Furthermore, the time dependence can be modeled directly. This reinforces the earlier claim that separation of estimation and decision making has beneficial consequences for applications of this nature.

Figure 8 shows a plot of the accuracy of the various models as a function of the sample size. For a fixed sample size k, $50 \leq k \leq 400$; for each model a subsample of size k was randomly selected from the total sample of size 450 from day1. For each k, this sampling was repeated 10 times, the classifiers were trained on the random sample, and their performance was evaluated on the independent data from day2 (all 450 samples). Each point on the graph shows the mean accuracy over 10 such runs. A fixed-size network architecture was maintained for each of the sample sizes since little variation in performance was evident from changing the number of units or layers. The fact that the estimates show little variation as the sample size varies is encouraging, since it shows that the classifier design technique is robust as a function of the amount of training data available. Note also that the time-domain classifier is consistently poorer than the others. Indeed, the ARX model outperforms the single large network for sample sizes greater than 250, indicating that the large

network may be overfitting the noise in the data due to its many degrees of freedom. The multiple-network architecture is consistently better than the other models.

E. Time-Averaged Classification Results

While the ARX model exploits local correlation information at the time-series level, it has already been mentioned that it would be desirable to incorporate correlation information on a more global scale, based on the belief that faults are more likely to persist from one 4-sec window to the next than they are to change.

A relatively simple delayed-decision component was implemented which takes the product of network outputs for each class over the current and some past number of outputs, i.e.,

$$p^n(c_i) = C \prod_{j=0}^{j=M} o_i^{n-j}$$

where $p^n(c_i)$ is the estimated probability of class i at time n, o_i^n is the network output at node i at time n, M is the "memory," and C is a constant independent of the class i. Normalized probability estimates can be obtained by setting

$$\hat{p}^n(c_i) = \frac{p^n(c_i)}{\sum_{k=1}^{m} p^n(c_k)}$$

where m is the number of classes. Appendix B gives a simple analysis of the behavior of this estimator, in particular its robustness to random errors with respect to making classification decisions. Not surprisingly, making the memory M longer reduces the probability of an incorrect decision or false alarm due to random noise effects. Figure 9 shows the smoothing effect of using time-correlated information in this manner on a test data set, using a memory of size 5. In the graph at the bottom, the time-correlation technique results in far less noise in the classification decision than in the other graphs where no memory is used. In the case of the two multiple networks described in the last section (using no information about time correlation), with 8.22-percent and 6.44-percent error rates, the post-processing to take time-correlation into account reduces the error to 2.33 percent and 1.78 percent, respectively.

There is a trade-off in increasing the memory M, in the sense that a longer memory will lead to longer delays in detecting a transition to a new fault. In fact, in Fig. 9, most of the errors occur just after fault-transition boundaries (windows 75, 150, 300, and 375) due to the effective lag in detecting change. More sophisticated decision strategies using context-dependent memory, Markov models, and information from non-neural detectors (such as monitoring the autoregression error sequence to detect change) are also being investigated.

Table 2 shows the actual error rates for the various architectures using a memory of size 5. There is a universal improvement in performance except for one run with the time-domain features, where the smoothing actually made the results worse as compared with no smoothing. The other models achieve error rates between 2 and 3 percent. Table 3 shows the confusion matrix corresponding to the 1.78-percent error rate for the first multiple network in Table 2. Some tachometer failures are misclassified as tachometer noise problems. Also there are some false alarms, two bursts of 8 sec each during the 15 min of normal data. On closer inspection, both of these 8-sec bursts occurred on the "fault" boundaries, when the true class changed from a non-normal to a normal fault, and as such are artifacts of the 20-sec memory scheme. Such false alarms during transitions are not nearly as serious as those that might occur during continuous normal conditions. No such false alarms occurred in the tests described here.

VII. Conclusion

This article describes an application of neural network techniques to pattern classification for a real-world fault-diagnosis task. In particular, the advantages of employing a modular architecture for this problem have been demonstrated, where domain knowledge is brought to bear in designing the lower-level signal-processing modules and the higher-level decision process, and where the task of mapping feature values to class labels is assigned to the neural network model. The initial model proposed was successfully validated on field data.

Acknowledgments

The authors thank D. Ginavan and D. Custer of the Bendix Field Engineering Corporation, and E. C. Posner and R. Stevens of the Office of Telecommunications and Data Acquisition for their assistance in obtaining summaries of Discrepancy Report data.

References

[1] A. S. Willsky, "A Survey of Design Methods for Failure Detection in Dynamic Systems," *Automatica*, vol. 12, pp. 601–611, 1976.

[2] P. Smyth, "Automated Monitor and Control for Deep Space Network Subsystems," *TDA Progress Report 42-98*, vol. April–June 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 110–120, August 15, 1989.

[3] P. Smyth and J. Mellstrom, "Initial Results on Fault Diagnosis of DSN Antenna Control Assemblies Using Pattern Recognition Techniques," *TDA Progress Report 42-101*, vol. January–March 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 136–151, May 15, 1990.

[4] W. Gawronski and J. Mellstrom, "Elevation Control System Model for the DSS 13 Antenna," *TDA Progress Report 42-105*, vol. January–March 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 83–108, February 15, 1991.

[5] J. Miller, R. Goodman, and P. Smyth, "Objective Functions for Probability Estimation," to appear in *Proceedings of the International Joint Conference on Neural Networks*, Seattle, 1991.

[6] H. Gish, "Maximum Likelihood Training of Neural Networks," to appear in *AI and Statistics 3: Proceedings of the 1991 Workshop*, D. Hand, ed., Chapman and Mall, London, forthcoming.

[7] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.

[8] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

[9] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing, Vol. 1*, Cambridge, Massachusetts: MIT Press, pp. 318–362, 1986.

[10] D. Touretzky, ed., *Advances in Neural Information Processing, vols. 1, 2, and 3*, San Mateo, California: Morgan Kaufman Publishers, 1989, 1990, and 1991.

[11] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Redwood City, California: Addison-Wesley, 1991.

[12] R. Hecht-Neilsen, *Neurocomputing*, Reading, Massachusetts: Addison-Wesley, 1990.

[13] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data," *IEEE Trans. Geoscience and Remote Sensing*, vol. 28, no. 4, pp. 540–552, July 1990.

[14] S. M. Weiss and I. Kapouleas, "An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine Learning Classification Methods," *Proceedings of IJCAI 1989*, Palo Alto, California, pp. 781–787, 1989.

[15] L. Ljung, *System Identification—Theory for the User*, Englewood Cliffs, New Jersey: Prentice-Hall, 1987.

[16] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Computation*, vol. 3, no. 1, 1991.

[17] I. Guyon, P. Albrecht, Y. LeCun, J. Denker, and W. Hubbard, "Design of a Neural Network Character Recognizer for a Touch Terminal," *Pattern Recognition*, vol. 24, no. 2, pp. 105–119, 1991.

[18] Y. Mori and K. Joe, "A Large-Scale Neural Network Which Recognizes Handwritten Kanji Characters," *Advances in Neural Information Processing, Vol. 2*, D. Touretzky, ed., San Mateo, California: Morgan Kaufman Publishers, pp. 415–422, 1990.

[19] Y. LeCun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard, "Handwritten Digit Recognition: Application of Neural Network Chips and Automatic Learning," *IEEE Communications Magazine*, vol. 27, no. 11, pp. 41–46, November 1989.

[20] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 328–339, March 1989.

Table 1. Classification rates on test data sets for classifiers with no time-correlation information

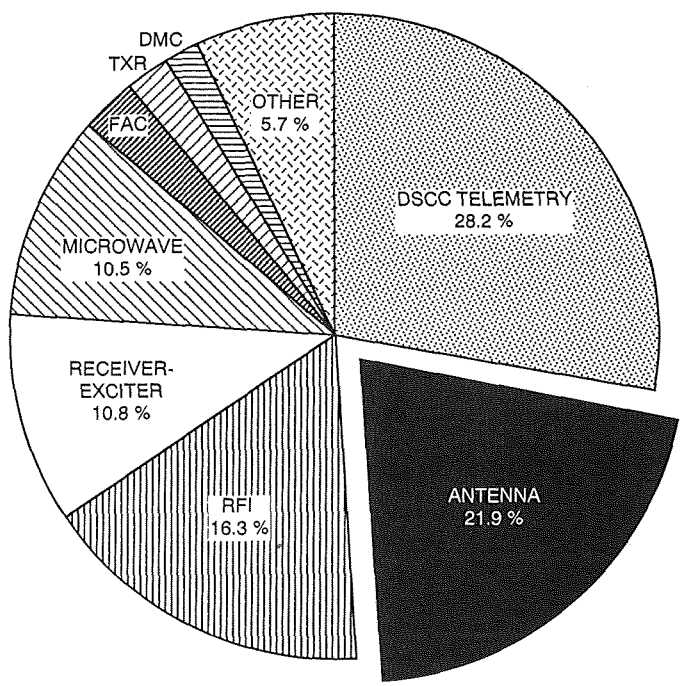
Type of network	Percentage correct		
	Test on day2	Test on day1	Mean
Multiple network	91.78	93.56	92.67
Single large network	90.22	94.00	92.11
ARX model features	91.11	91.78	91.45
Time-domain features	80.00	86.67	83.34

Table 2. Classification rates on test data sets for classifiers using time-correlation information, with memory set to 5 (20 sec)

Type of network	Percentage error		
	Test on day2	Test on day1	Mean
Multiple network	1.78	2.33	2.06
Single large network	3.33	2.33	2.83
ARX model features	3.22	2.33	2.78
Time-domain features	33.44	4.22	18.83

Table 3. Confusion matrix obtained when testing on day1 data set using time-correlation information, with memory set to 5 (20 sec)

True class	Estimated class			
	Tachometer noise	Tachometer failure	Compensation loss	Normal
Tachometer noise	75	0	0	0
Tachometer failure	6	69	0	0
Compensation loss	0	0	73	2
Normal	0	2	2	221



DMC: DSCC MONITOR AND CONTROL
TXR: TRANSMITTER
FAC: DSCC TECHNICAL FACILITIES
RFI: RADIO FREQUENCY INTERFERENCE
DSCC: DEEP SPACE COMMUNICATIONS COMPLEX

Fig. 1. Telemetry data outages during scheduled tracks, January 1, 1986 to July 31, 1990, by DSN subsystem, expressed as a percentage of total hours lost. (Total hours lost was 3.08% of the scheduled support time.)

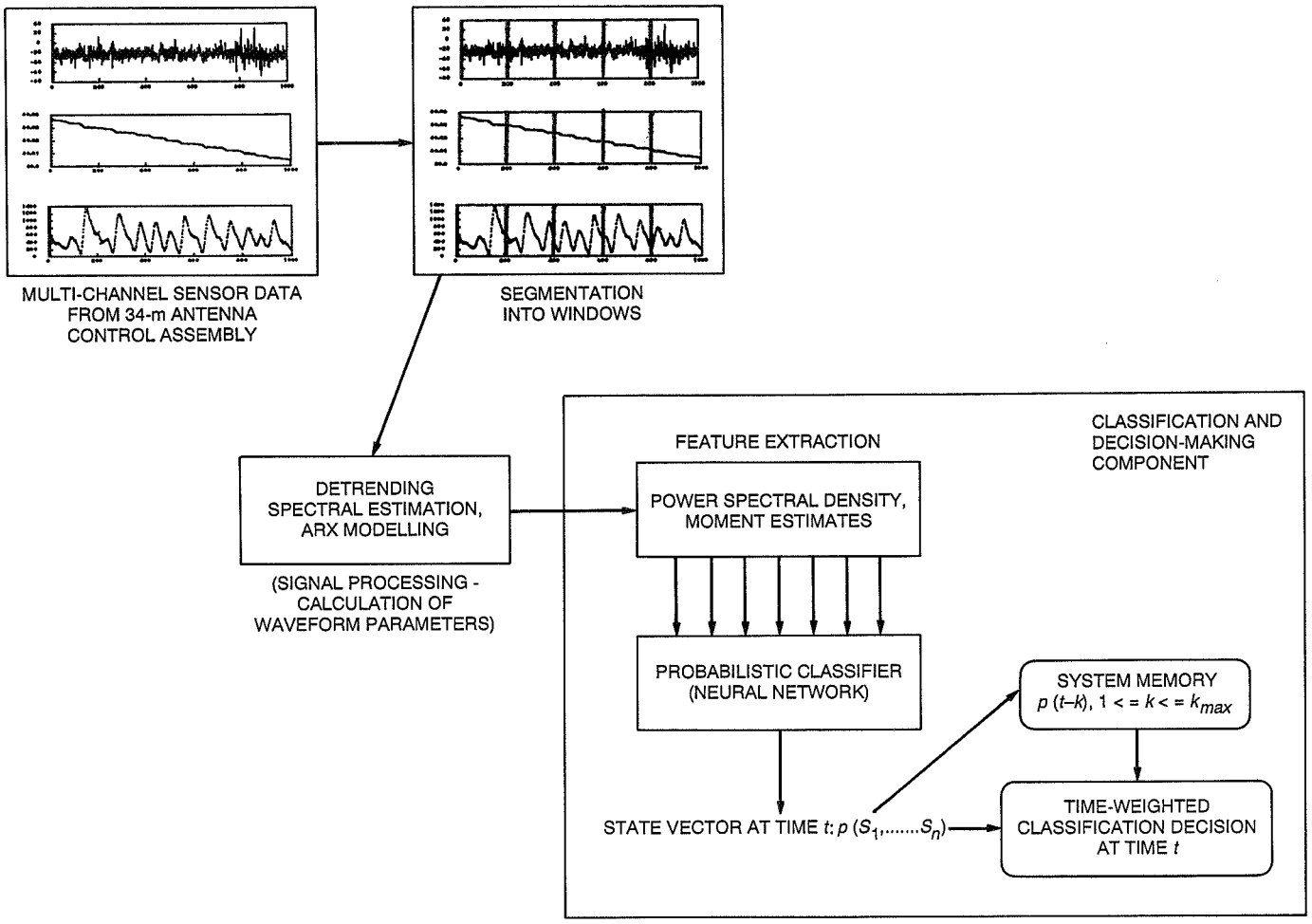


Fig. 2. Hierarchical information flow of the system.

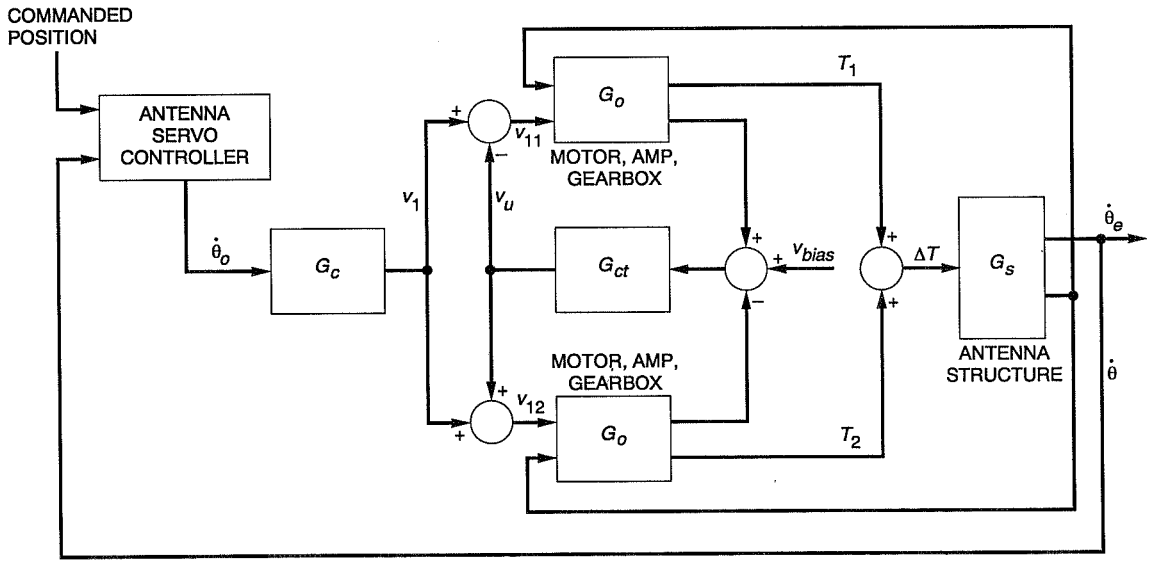


Fig. 3. The 34-m antenna elevation axis drive assembly.

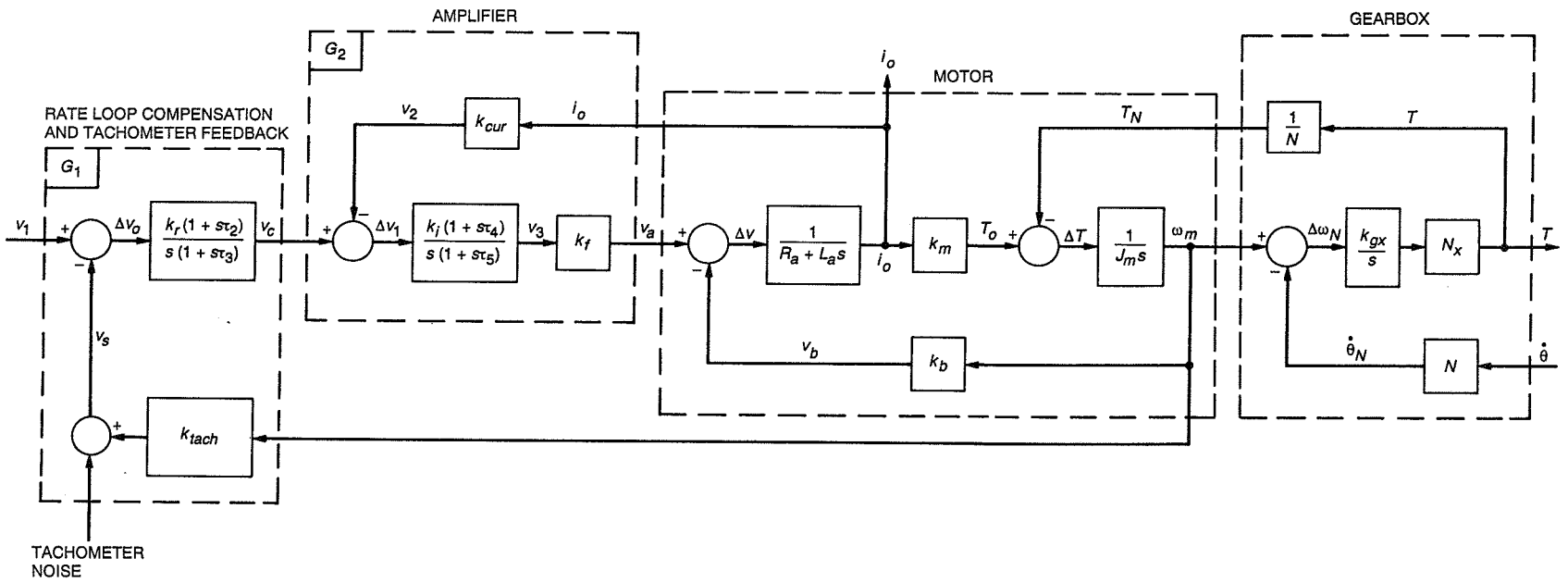


Fig. 4. Motor, amplifier, and gearbox set.

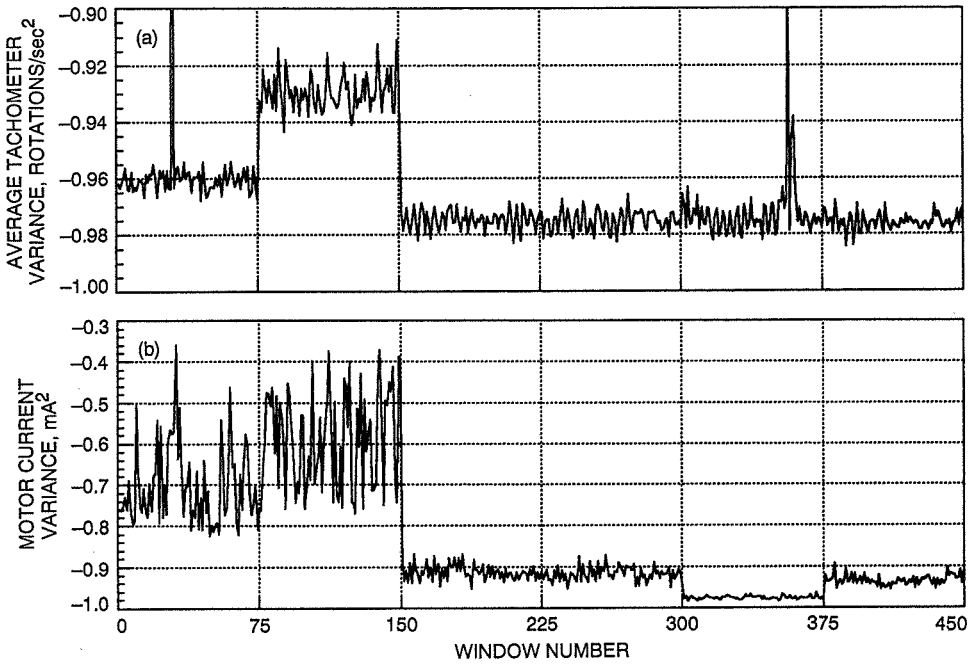


Fig. 5. Discrimination capability of (a) variance of the average tachometer sensor, and (b) motor current variance (day2 data).

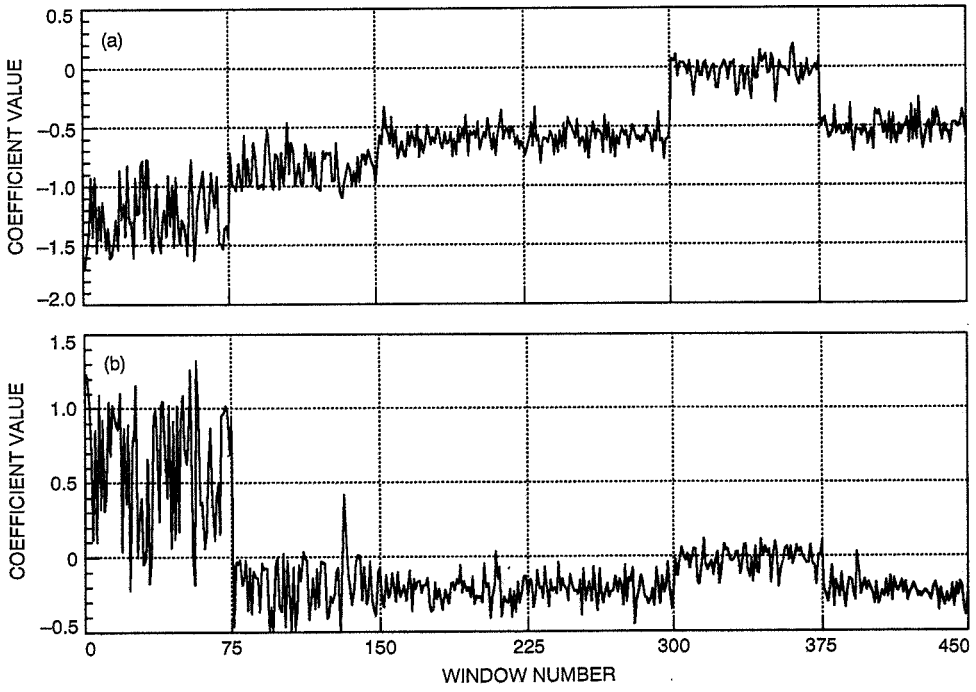


Fig. 6. Discrimination capabilities of (a) ARX coefficient 1, and (b) ARX coefficient 2 (day2 data).

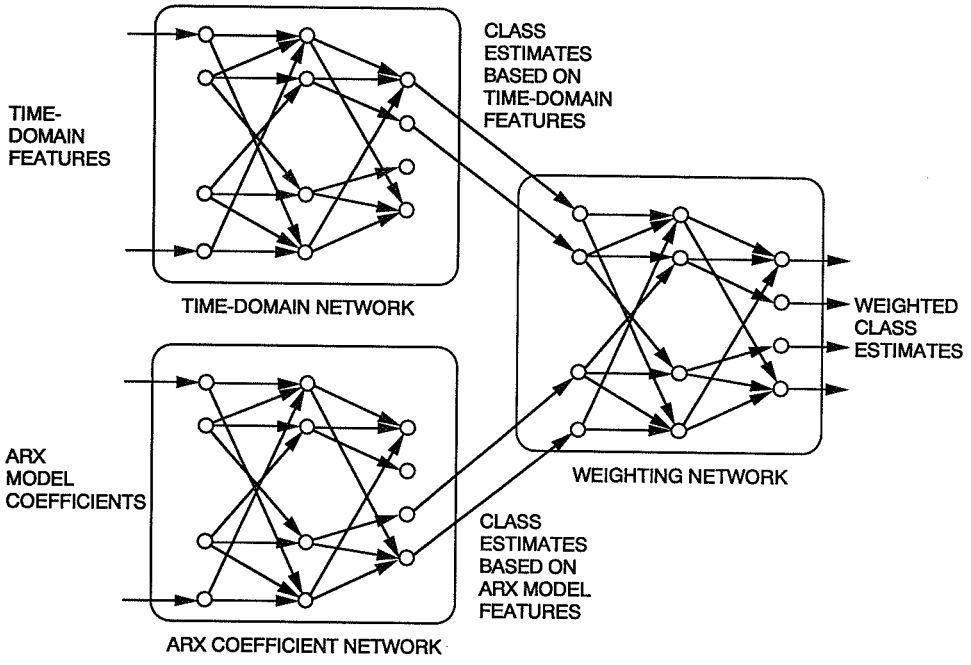


Fig. 7. Multiple-network architecture.

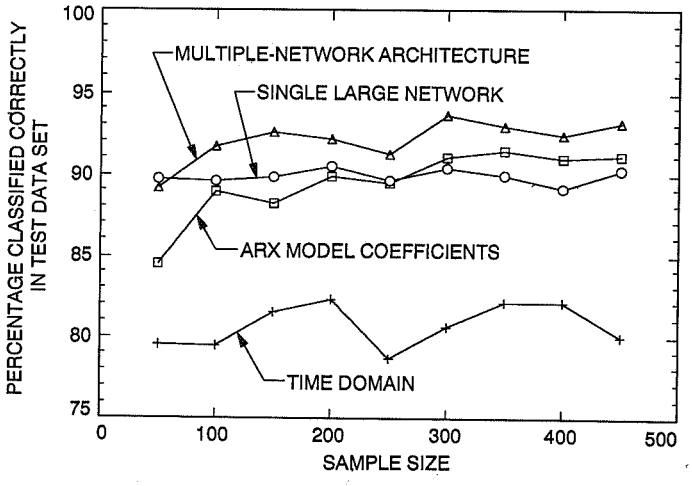


Fig. 8. Classification accuracy of classifier models as a function of sample size.

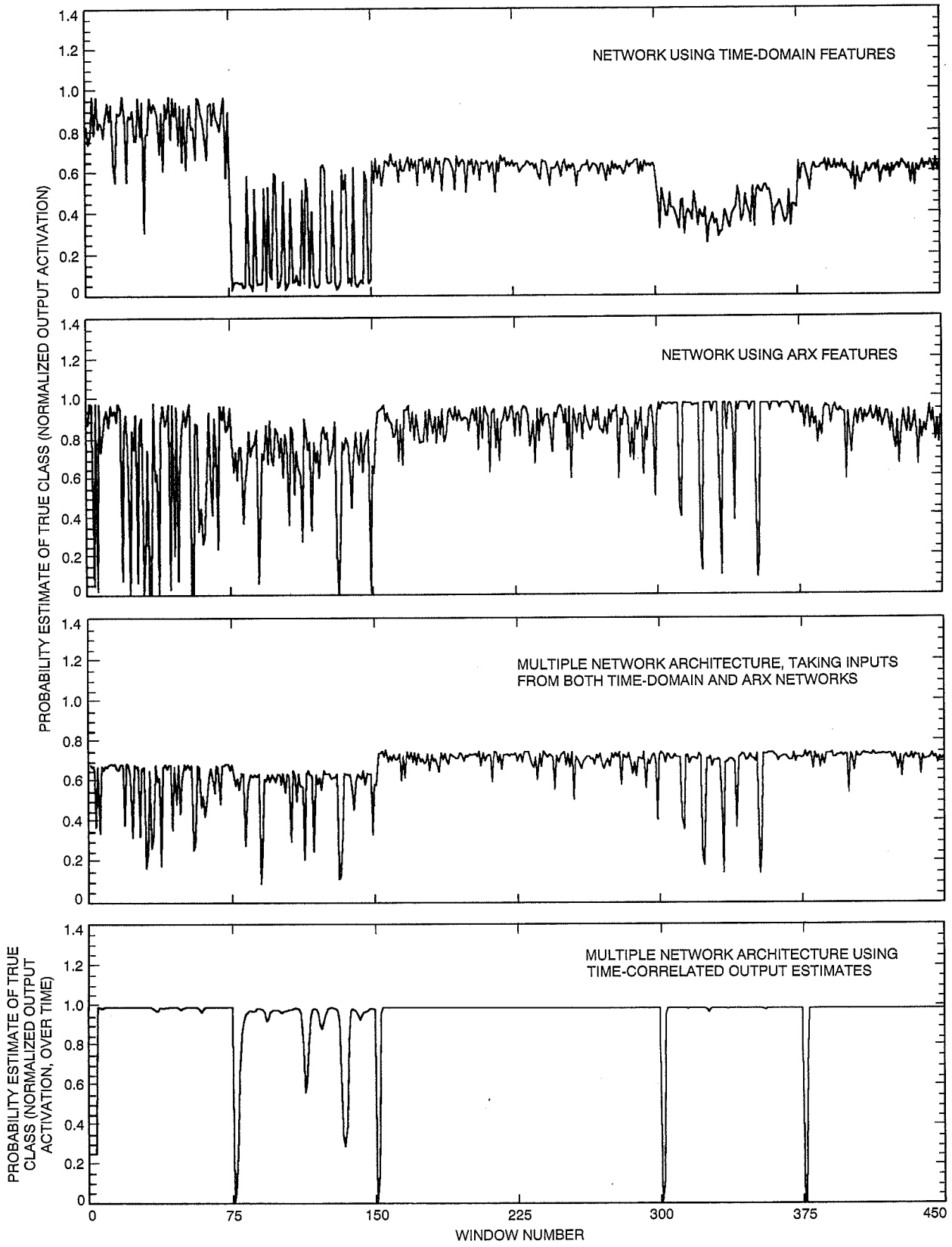


Fig. 9. Smoothing effect of using time-correlated information on a test data set (memory of size 5).

Appendix A
Neural Network Model Description

The following description of an example of a popular feed-forward multilayer neural network model will familiarize the reader with the general notation and concepts. This appendix is essentially a repetition of the Appendix in [3] and is included in order to make the report self-contained.

Figure A-1 shows an example of a network. The input nodes are labeled n_i, $1 \leq i \leq K+1$; the hidden nodes are labeled h_j, $1 \leq j \leq H$; and the output layers are labeled o_k, $1 \leq k \leq m$. In general, there are $K+1$ input units, where K is the number of features. The extra node is always in the "on" state, providing a threshold capability. Similarly, there are m output nodes, where m is the number of classes.

The number of hidden units H in the model is chosen based on rules of thumb and empirical experience. The size of this hidden layer can influence the classifier performance in the following manner: Too many hidden units, and the network overfits the data (i.e., the estimation error will be large); too few hidden units, and the network is left with insufficient representational power (i.e., the approximation error term is large).

Each input unit i is connected to each hidden unit j by a link with weight w_{ij}, and each hidden unit j is connected to each output unit k by a weighted link w_{jk}. Each hidden unit calculates a weighted sum and passes the result through a nonlinear function $F()$, i.e.,

$$a(h_j) = F\left(\sum_{i=1}^{i=K+1} w_{ij}\, a(n_i)\right)$$

where $a(n_i)$ is the activation of input unit i. Typically, this is just a linear (scaled) function of the input feature. A commonly used nonlinear function in the hidden unit nodes $F(x)$ is the so-called sigmoid function, defined as

$$F(x) = \frac{1}{1+e^{-x}}$$

Output unit k calculates a similar weighted sum using the weights w_{jk} between the jth hidden unit and the kth output unit, i.e.,

$$a_k = G\left(\sum_j w_{jk}\, a(h_j)\right)$$

where a_k is the activation of the kth output node. The function $G(x)$ can be chosen as either a linear (e.g., $G(x) = x$) or a nonlinear function. For example, for a classification problem such as that described in this article, the sigmoid function is used to restrict the range of the output activations to the range $[0, 1]$.

A classification decision is made by choosing the output unit with the largest activation for a given set of inputs (feature values); i.e., choose class k such that

$$k = \arg\max_i \{a_i\}$$

The network design problem is then to find the best set of weights such that a particular objective function is minimized on the N training data samples. The training data are in the form of input-output pairs $\{\underline{x}_i, y_i\}$, $1 \leq i \leq N$, where $\underline{x}_i$ is a feature vector and y_i is the desired output (for simplicity of notation, assume that there is just a single output model). Let $\hat{y}_i(\Omega, \underline{x}_i)$ be the network output for a particular set of weights Ω and input vector $\underline{x}_i$. The objective function is typically some metric on y_i and $\hat{y}_i$, whose mean value is estimated on the training data. Such commonly used objective functions include the mean-squared error

$$E_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i(\Omega, \underline{x}_i)\right)^2$$

and the cross-entropy error

$$E_{CE} = \frac{1}{N}\sum_{i=1}^{N} y_i \log \frac{y_i}{\hat{y}_i(\Omega, \underline{x}_i)} + (1-y_i)\log\frac{1-y_i}{1-\hat{y}_i(\Omega, \underline{x}_i)}$$

From a maximum-likelihood standpoint, the mean-squared-error approach assumes that the training data are perturbed by additive Gaussian noise, while the cross-entropy function assumes a multinomial distribution on

the class labels. Despite these differences, for classification problems there appears to be little significant difference between these objective functions; in fact, as shown in [5], these two error functions are the simplest functions in the general family of functions that can be proven to asymptotically produce consistent probability estimates. For the experiments reported in this article the mean-squared error objective function was used.

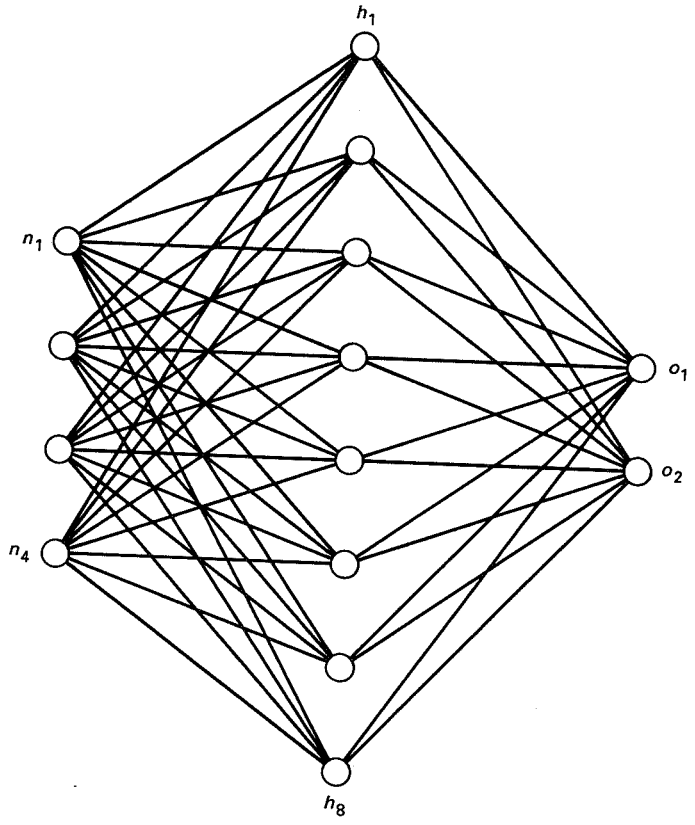


Fig. A-1. Example of a three-layer feed-forward neural network.

Appendix B

Error Probability as a Function of Memory Length

A simple model is proposed to determine the effects of memory length on probability of error when using the product of output activations over time as a classification metric. Let n be the number of terms in the product and $n = M + 1$, where M is the memory length referred to in the text. Assume a 2-class problem: this is a worst case since the multiclass case can be reduced to a 2-class problem where the noise is malicious in the sense that all misclassification errors occur in the form of one other particular class. Let o_1 be the classifier's output when class 1 is actually the true class, and let $\hat{o}_1$ be the output of the classifier when a noise hit occurs, i.e., the output for class 2 is larger than for class 1, even though class 1 is the true class. Such noise hits are likely, given the noisy environment in which the sensors operate. Outputs o_2 and $\hat{o}_2$ are defined in a similar manner; for example, one might have $o_1 = 0.8$, $o_2 = 0.1$, $\hat{o}_1 = 0.4$, and $\hat{o}_2 = 0.5$. Typically the network produces output activations where it is much more confident in its correct decisions than it is in its mistakes. Note that the simplifying assumption has been made that the outputs o_1, etc., are constant with respect to time. This makes the model simpler to analyze and gives a general idea of what is happening. A more sophisticated approach might be to assume that o_1 is the mean value of the activation which varies with some deviation σ_1.

In the case where class 1 is the true class, for a product of size n, if k noise hits occur, then the probability estimate for class 1 will only be greater than class 2 if

$$\sum_{n-k} \log \frac{1}{o_1} + \sum_{k} \log \frac{1}{\hat{o}_1} > \sum_{n-k} \log \frac{1}{o_2} + \sum_{k} \log \frac{1}{\hat{o}_2}$$

or

$$n > k(1+\gamma)$$

where

$$\gamma = \log \frac{\hat{o}_2}{\hat{o}_1} / \log \frac{o_1}{o_2}$$

Hence the "error-correction" capability of a memory-based classifier depends on this parameter γ, which typically is quite small, perhaps about 0.1 in many cases, i.e., the classifier is 10 times more confident in its decisions when it is correct than when it is incorrect. A value of $\gamma = 1.0$ means that there is no difference, and the error-correction capability is effectively reduced to $n/2$ for a product of size n.

Now let p be the probability of such a noise hit. This parameter p measures the reliability of the classifier and the robustness of the features to withstand noise in the data. Assuming that these noise hits are not correlated (which may not be a realistic assumption in practice, but the model should be kept simple), then the error probability is given by the total number of ways in which at least $n/(1+\gamma)$ errors can occur in a window of size n. This is simply the sum of binomial terms:

$$p_e = \sum_{i=\lceil \frac{n}{1+\gamma} \rceil}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

If $\gamma << 1$, $\lceil \frac{n}{1+\gamma} \rceil \approx n$ so that

$$p_e \approx p^n$$

i.e., the error falls off exponentially as a function of memory.

A Cooled 1- to 2-GHz Balanced HEMT Amplifier[1]

G. G. Ortiz
Radio Frequency and Microwave Subsystems Section

S. Padin
Caltech Owens Valley Radio Observatory
Big Pine, California

The design details and measurement results for a cooled L-band (1 to 2 GHz) balanced high-electron-mobility-transistor (HEMT) amplifier are presented here. The amplifier uses commercially available packaged HEMT devices (Fujitsu FHR02FH). At a physical temperature of 12 K the amplifier achieves noise temperatures between 3 and 6 K over the 1- to 2-GHz band. The associated gain is approximately 20 dB.

I. Introduction

Cryogenically cooled high-electron-mobility-transistor (HEMT) amplifiers have realized noise temperatures as low as the operating frequency (in GHz) of the amplifier up to 43 GHz [1,2]. These amplifiers have now become the standard for radio astronomy applications. Cooled HEMT amplifiers are also used as the first intermediate frequency (IF) stage in millimeter-wave superconductor-insulator-superconductor (SIS) and Schottky mixer systems. In these receivers the noise performance of the IF amplifier is very important because the mixer is usually lossy.

Most millimeter-wave receiver systems built for radio astronomy use an L-band (1 to 2 GHz) IF amplifier with a bandwidth of approximately 500 MHz and a noise temperature of approximately 4 K [3,4]. In a typical 115-GHz SIS receiver the IF amplifier contributes approximately 30 percent of the total receiver noise. A bandwidth of 500 MHz is barely sufficient for observations of sources with high velocity dispersion and it limits the capability of systems to observe several molecular transitions simultaneously. Increasing the receiver bandwidth by using a higher frequency IF amplifier is not acceptable because this would increase the IF noise contribution and therefore degrade the sensitivity of the system. The authors' approach to this problem has been to develop a 1-GHz bandwidth L-band cooled HEMT amplifier, while retaining < 6-K noise temperature.

A significant problem in the design of a wideband cooled amplifier is obtaining scattering parameters (s-parameters) and noise parameters for transistors at cryogenic temperatures. For this work, no facilities were available for measuring low-temperature s-parameters, but the HEMT noise parameters were measured at a physical temperature of 12 K. The absence of s-parameter information precluded the design of a feedback amplifier, so a bal-

[1] This project was partly supported by the Caltech President's Fund.

anced configuration was adopted. This has the advantage of providing a good input match even though the amplifiers in the two arms of the balanced circuit are poorly matched. However, there are disadvantages. The loss of the input hybrid degrades the noise temperature, and coupling errors in the hybrids and differences between the amplifiers reduce the gain and result in a noise contribution from the input load. In the amplifier described here these effects degrade the noise temperature by less than 1 K.

II. Noise in a Balanced Amplifier

The noise contributions in a balanced amplifier are explored in Fig. 1. Each hybrid directs a fraction c of the input power to the 0-deg port and the remaining power to the 90-deg port. The deviation from quadrature at the outputs is θ. To simplify the analysis, the amplifiers are assumed to have similar gains but different transfer function phases. In practice, this situation can be approached by selecting similar devices and by adjusting the bias.

The power gain of the balanced amplifier (with the input terminated in a matched source and the output terminated in a matched load) is

$$G = 2gc(1-c)(1+\cos\phi) \tag{1}$$

where g is the power gain of each amplifier in the balanced structure and ϕ is the phase difference between the amplifier transfer functions. The output noise temperature with the input terminated in a matched source at 0 K is

$$T_{out} = gT_o\left[(2c^2-2c+1)-2c(1-c)\cos\phi\right]+gT_a \tag{2}$$

where T_o is the physical temperature of the input hybrid termination and T_a is the input equivalent noise temperature of each amplifier. The first term is the contribution from the input hybrid termination and the second term represents the noise generated by the amplifiers. Note that the noise from the two amplifiers is uncorrelated. The input noise temperature of the balanced amplifier is

$$T_n = \frac{T_{out}}{G}$$

$$T_n = \frac{T_o[(2c^2-2c+1)-2c(1-c)\cos\phi]+T_a}{2c(1-c)(1+\cos\phi)} \tag{3}$$

The loss of the input hybrid can be modeled as an attenuator at the amplifier input. Loss in the output hybrid affects only the overall gain (and hence the noise contribution of the next stage). With an input hybrid loss L, the noise temperature of the balanced amplifier is

$$T_n' = T_o(L-1)+LT_n \tag{4}$$

where T_n is given by Eq. (3). As an example of what might be achieved, a balanced circuit containing amplifiers with $T_a = 4$ K, $\phi = 5$ deg, hybrids with coupling errors of 1 dB ($c = 0.40$), and excess loss of 0.1 dB would have a noise temperature of 5.09 K at a physical temperature of 12 K. If the input load were at 4 K instead of 12 K, the noise temperature would be 4.73 K.

III. HEMT Device Noise Parameters

Device noise parameters for this work were obtained from measurements of the noise temperatures of several single-ended amplifiers each with a different input-matching network. The same HEMT device was used for the entire set of measurements. Previous work with GaAs field effect transistors (FETs) at L-band provided an estimate of the optimum source impedance so that it was necessary to explore only a small part of the source impedance plane in order to determine the HEMT noise parameters [5].

Different source impedances were provided using a cryogenically coolable test fixture. The input section consisted of a microstrip transformer realized on 60-mil thick RT/Duroid 6002[2] ($e_r = 2.94$) and a series chip inductor[3] close to the HEMT package. The output section was a 50-Ω microstrip line of the same substrate type. Connections to the HEMT device were simple pressure contacts which allowed easy removal of the input-matching network assembly. The input substrates were transferred to a different fixture fitted with sub-miniature series A (SMA) connectors to allow measurement of the source impedance presented to the HEMT.

Noise temperature measurements were made using the arrangement of Fig. 2. A cooled 20-dB attenuator at the amplifier input provides a cold load, and signals enter and leave the cryostat via low-loss coaxial lines. The losses of the lines, connectors, input bias network, and the 20-dB attenuator were measured at 12 K. These data along with measurements of the noise temperature of the receiver in

[2] RT/Duroid 6002 manufactured by Rogers Corporation, Microwave Materials Division, 100 S. Roosevelt Ave., Chandler, Arizona.

[3] Surface mount inductor series 1008CS manufactured by Coilcraft, 1102 Silver Lake Rd., Cary, Illinois.

Fig. 2 were used to calculate the noise temperature of the amplifier from measurements of the system noise temperature at the cryostat input. A solid-state noise source calibrated against liquid nitrogen and ambient loads was used for the measurements.

The amplifier noise temperature is related to the source impedance, $R + jX$, by

$$T = T_{min} + 290\frac{g_n}{R}\left[(R - R_{opt})^2 + (X - X_{opt})^2\right] \quad (5)$$

where $R_{opt} + jX_{opt}$ is the optimum source impedance, g_n is the noise conductance, and T_{min} is the minimum noise temperature [6]. The noise parameters $\{R_{opt}, X_{opt}, g_n, T_{min}\}$ were obtained from the results of ten measurements of T with different source impedances using a least-squares fit to Eq. (5). Although the final amplifier was constructed using Fujitsu FHR02FH (200 μm × 0.25 μm) devices, the noise parameters were measured for a General Electric (GE) (300 μm × 0.25 μm) AlGaAs/GaAs HEMT in a standard 70-mil package [7]. Recent work with FHR02FH chips indicates that the packaged GE and packaged Fujitsu devices have very similar R_{opt} and X_{opt} values at L-band [8]. The measured noise parameters for the GE HEMT are summarized in Table 1. FHR02FH devices have g_n values approximately equal to 0.03 milliSiemens (mS) and are therefore less sensitive than the GE device to deviations from the optimum source impedance.

IV. Amplifier Design

The input-matching circuit for each arm of the balanced amplifier consists of a 71-Ω line ($\lambda/4$ at 1.5 GHz) and a 10-nH series inductor. This gives a good noise match over the 1- to 2-GHz band and is reasonably compact and low loss. The impedance of the transformer and the value of the inductor were chosen to minimize $\left[(R - R_{opt})^2 + (X - X_{opt})^2\right]$ over the 1- to 2-GHz band. Figure 3 shows measurements of the impedance presented by the matching circuit along with the optimum source impedance from Table 1. The calculated increase in noise temperature due to deviation of the source impedance from optimum is approximately 1 K. A single-ended amplifier was constructed with the input-matching circuit described above and a 50-Ω load at the output. The performance is shown in Fig. 4. As expected, the noise temperature increases by approximately 1 K at the band edges. Since the gain shows only a gentle slope across the band, no attempt was made to adjust the output match.

The 90-deg hybrids in the balanced amplifier are 6-finger Lange couplers [9]. These are realized on the same 60-mil RT/Duroid 6002 substrate as used for the matching networks. At room temperature, the measured coupler insertion loss is approximately 0.2 dB, the coupling error is < 0.5 dB, and the deviation from phase quadrature is < 2 deg.

A schematic of the complete balanced amplifier is shown in Fig. 5 and the construction details are indicated in Fig. 6. To ensure good thermal coupling to the box, the HEMT devices are clamped to posts which protrude through the circuit board. The bias circuits are configured to allow independent biasing of the two HEMTs, although in practice this has not been necessary.

V. Results

Two prototype balanced amplifiers were constructed with Fujitsu FRH02FH HEMTs, and the noise temperatures and gains at a physical temperature of 12 K are shown in Figs. 7 and 8. Both amplifiers have noise temperatures in the range of 3 to 6 K. For these measurements the bias of both HEMT devices was adjusted to minimize the noise temperature. In Figs. 7 and 8, both transistors have the same bias. No other tuning was done, so the results indicate typical performance for a production amplifier. Some improvement in noise temperature might be obtained by adjusting the values of the inductors in the matching networks.

Input and output return losses were measured for one amplifier at 12 K and these data are shown in Fig. 9. The worst-case input return loss is 13 dB, but over most of the band the return loss is better than 15 dB. All the results were obtained with no illumination of the HEMTs.

VI. Conclusions

This article was a description of a 1- to 2-GHz cooled balanced HEMT amplifier. At a physical temperature of 12 K the amplifier has a noise temperature in the range of 3 to 6 K and a gain of approximately 20 dB. The amplifier was designed primarily as a wideband IF amplifier for millimeter-wave radio astronomy, but it also has applications in wideband L-band receiver systems.

References

[1] S. Weinreb, R. Harris, and M. Rothman, "Millimeter-Wave Noise Parameters of High Performance HEMTs at 300 K and 17 K," *IEEE MTT-S International Microwave Symposium Dig.*, pp. 813–816, 1989.

[2] K. H. Duh, W. F. Kopp, P. Ho, P. C. Chao, M. Y. Ko, P. M. Smith, J. M. Ballingall, J. J. Bautista, and G. G. Ortiz, "32 GHz Cryogenically Cooled HEMT Low-Noise Amplifiers," *IEEE Trans. Electron Devices*, vol. ED-36, pp. 1528–1535, August 1989.

[3] T. H. Buttgenbach, R. E. Miller, M. J. Wengler, D. M. Watson, and T. G. Phillips, "A Broad-Band Low-Noise SIS Receiver for Submillimeter Astronomy," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-36, pp. 1720–1726, December 1988.

[4] S. Weinreb, "SIS Mixer to HEMT Amplifier Optimum Coupling Network," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp. 1067–1069, November 1987.

[5] S. Weinreb, D. Fenstermacher, and R. Harris, "Ultra Low-Noise 1.2–1.7 GHz Cooled GaAs FET Amplifiers," NRAO Electronics Division Internal Report No. 220, National Radio Astronomy Observatory, Charlottesville, Virginia, 1981.

[6] H. A. Haus, ed., "IRE Standards on Methods of Measuring Noise in Linear Two Ports," *Proceedings of the IRE*, vol. 48, pp. 60–68, January 1960.

[7] K. H. Duh, M. W. Pospieszalski, W. F. Kopp, P. Ho, A. A. Jabra, P. C. Chao, P. M. Smith, L. F. Lester, J. M. Ballingall, and S. Weinreb, "Ultra-Low-Noise Cryogenic High Electron Mobility Transistors," *IEEE Trans. Electron Devices*, vol. ED-35, pp. 249–256, March 1988.

[8] J. D. Gallego and M. W. Pospieszalski, "Design and Performance of Cryogenically-Coolable Ultra Low-Noise L-Band Amplifier," NRAO Electronics Division Internal Report No. 286, National Radio Astronomy Observatory, Charlottesville, Virginia, 1990.

[9] J. Lange, "Interdigitated Stripline Quadrature Hybrid," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 1150–1151, December 1969.

Table 1. Noise parameters for the GE HEMT

Frequency, GHz	R_{opt}, Ω	X_{opt}, Ω	g_n, mS	T_{min}, K
1.0	100 ± 20	160 ± 20	0.07 ± 0.03	4.5 ± 1
1.5	80 ± 10	160 ± 20	0.07 ± 0.03	3.5 ± 1
2.0	80 ± 20	120 ± 20	0.13 ± 0.09	4.5 ± 1

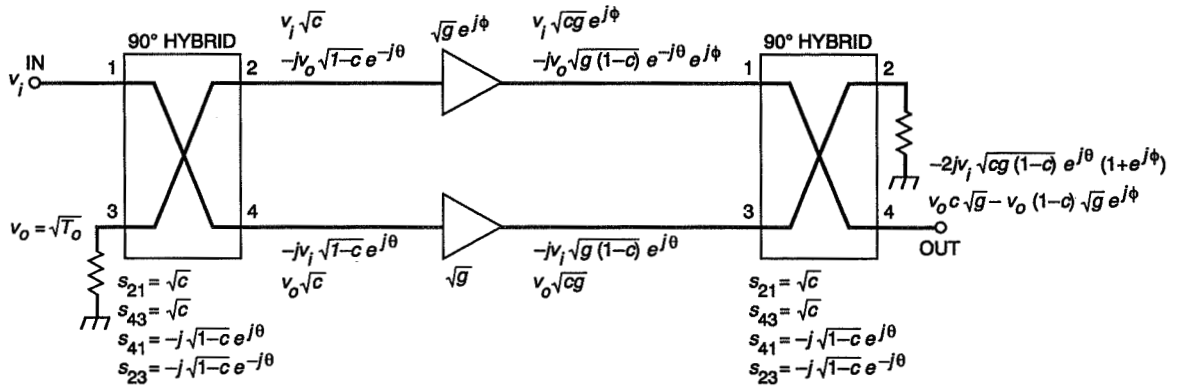


Fig. 1. Voltages in a balanced amplifier. The italicized terms indicate voltages due to input v_i. The other terms indicate noise voltages due to the input hybrid termination. The time dependent and delay terms have been omitted.

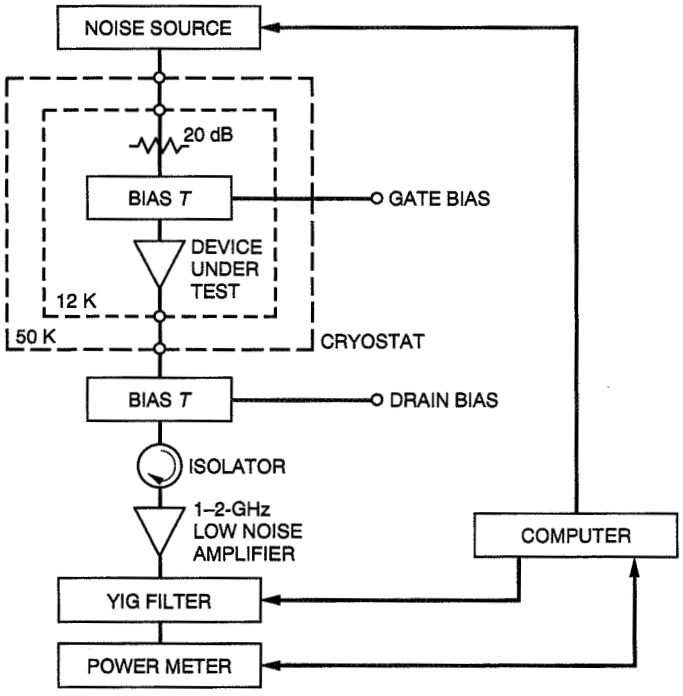


Fig. 2. Noise temperature measurement system.

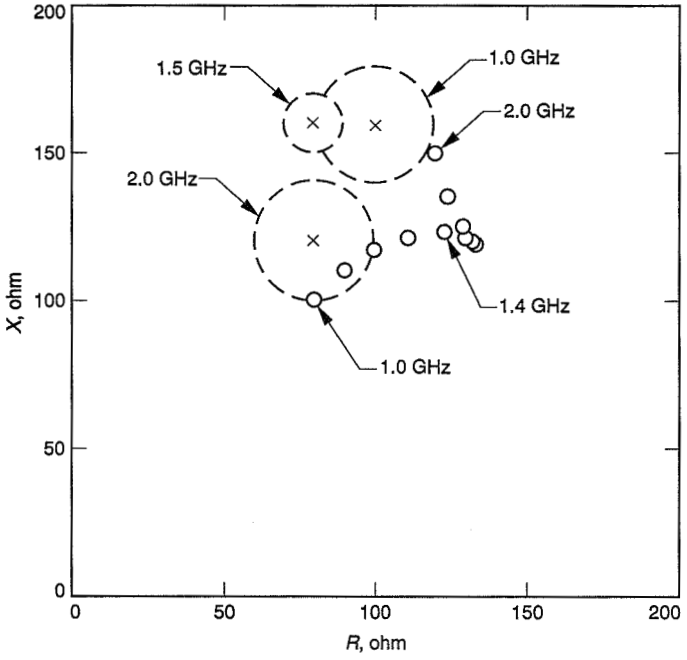


Fig. 3. Measured impedance of the amplifier input-matching circuit at 100-MHz intervals (solid circles) and the optimum source impedance from Table 1 (dashed circles). The size of the dashed circles indicates the uncertainty in the optimum source impedance measurements.

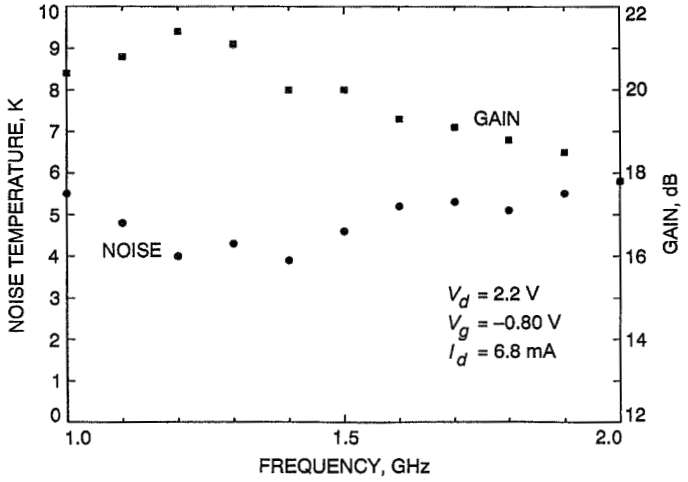


Fig. 4. Performance of a single-ended amplifier with a GE HEMT. The noise temperature uncertainty is ±1.5 K, where V_d = voltage at drain, V_g = voltage at gate, and I_d = current at drain.

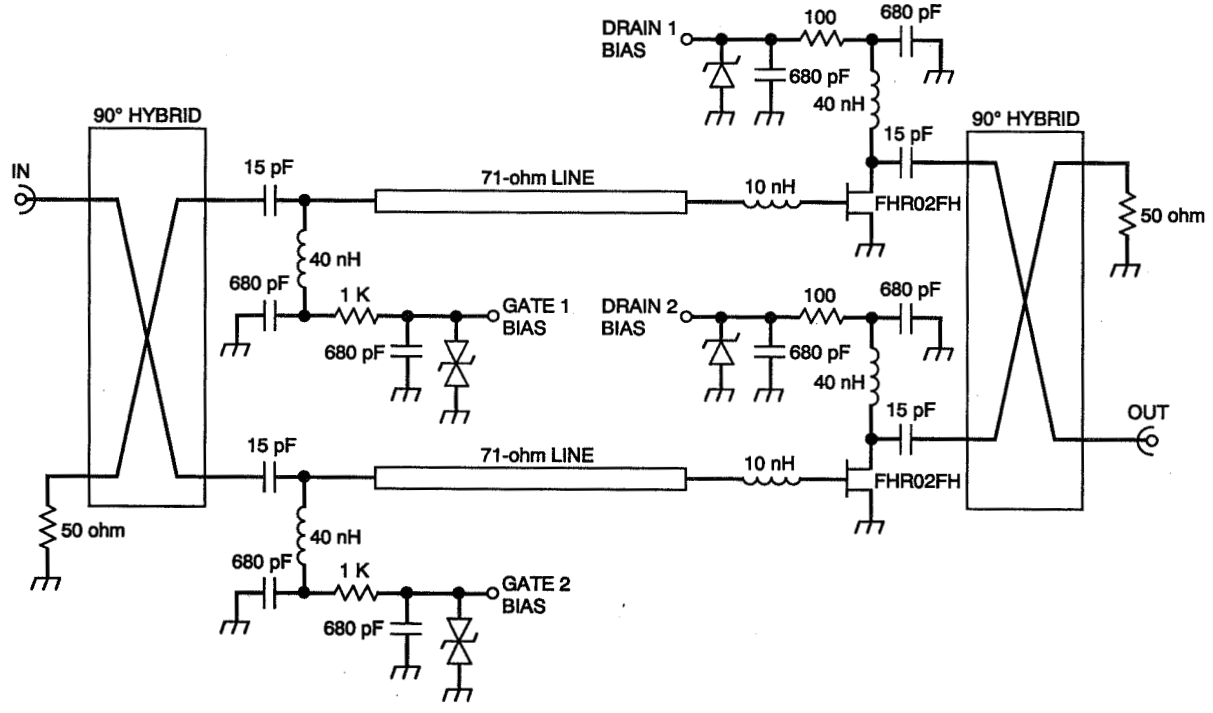


Fig. 5. Schematic of the 1- to 2-GHz balanced amplifier.

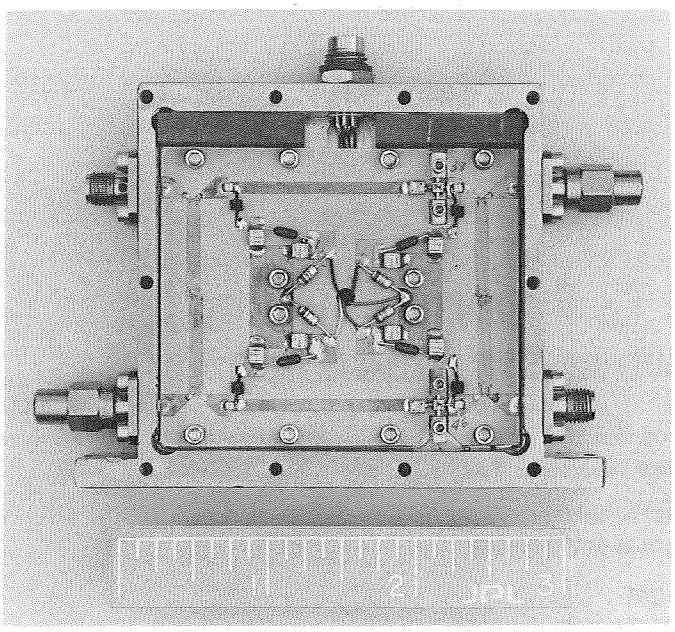


Fig. 6. Photograph of the 1- to 2-GHz balanced amplifier. (Scale shown in inches.)

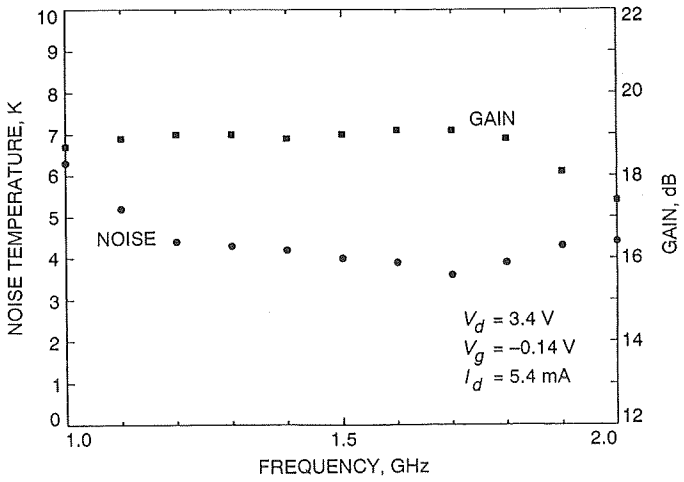


Fig. 7. Gain and noise temperature of balanced amplifier serial number 00 at a physical temperature of 12 K. The noise temperature measurement uncertainty is ±1.2 K.

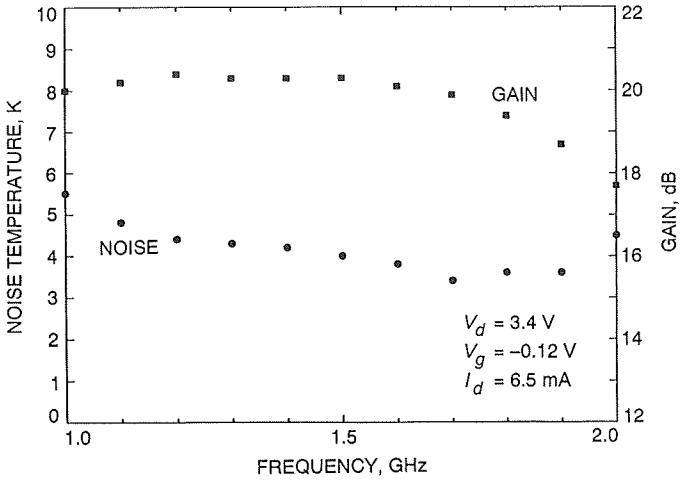


Fig. 8. Gain and noise temperature of balanced amplifier serial number 01 at a physical temperature of 12 K. The noise temperature measurement uncertainty is ±1.2 K.

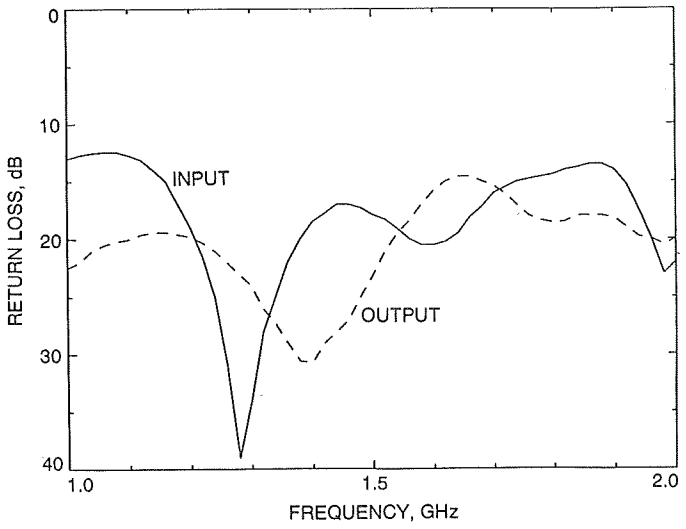


Fig. 9. Input and output return loss for amplifier serial number 00 at a physical temperature of 12 K.

Performance of a High T_c Superconducting Ultra-Low Loss Microwave Stripline Filter

J. J. Bautista and G. Ortiz
Radio Frequency and Microwave Subsystems Section

C. Zahopoulos and S. Sridhar
Department of Physics, Northeastern University

M. Lanagan
Argonne National Laboratory

This article reports successful fabrication of a five-pole interdigital stripline filter made of the 93-K superconductor ($Y_1Ba_2Cu_3O_y$) coated on a silver substrate, with center frequency of 8.5 GHz and an extremely high rejection ratio of 80 dB. The lowest insertion loss measured was 0.1 dB at 12 K, with a return loss of better than 16 dB, representing a significant improvement over a similar copper filter, and is comparable to low critical-temperature filters. The insertion loss appears to be limited by extrinsic factors, such as tuning mismatch and joint losses, and not by the superconducting material losses.

Recent results of the intrinsic microwave properties of high-quality single crystals [1] and thin films [2] of the high critical-temperature (T_c) superconductors raise tremendous prospects for the applications of these materials in microwave devices. These results of the surface resistance imply significantly improved performance as compared with devices using conventional materials, such as copper. It is clear, however, that significant problems need to be solved in order to translate the results on small scale samples into realistic structures necessary for actual devices.

A bandpass filter is of great utility in systems limited in performance by radio frequency interference (RFI) at the input. This is particularly important in space communications, where the front-end amplifier is a delicate high-electron mobility transistor (HEMT) or maser amplifier with a very low noise temperature. The effect of incident RFI is primarily determined by the level and frequency of the RFI. Both in-band and out-of-band RFI can result in significant gain compression, noise temperature increase, and, in the case of the maser, spurious output signals. For example, a 0.1-dB increase in insertion loss can result in a 0.4-K increase in noise temperature, which can be serious in low noise systems. Therefore, a usable (out-of-band) RFI filter should have stringent specifications: a narrow bandpass, a high out-of-band rejection ratio (at least 80 dB), and an extremely low insertion loss (~0.1 dB) to avoid in-band signal attenuation and added noise.

This article describes the fabrication and performance of an 8.5-GHz bandpass filter made of silver and coated with $Y_1Ba_2Cu_3O_{7-x}$ (YBCO). An interdigital tunable stripline resonator structure was selected over a microstrip filter as the optimal design due to its favorable performance characteristics [3], and its potential ability to meet the design criteria specified above. This structure is compact and, with the exception of waveguide filters, it has the highest unloaded resonator (Q) among the commonly used structure [4]. The filter, shown in Fig. 1, consists of five transverse electromagnetic (TEM) mode stripline resonators. Each resonator is approximately one-quarter wavelength long at the midband frequency and is short-circuited at one end and open-circuited at the other. These resonators are placed between two ground plates which are attached to the filter body by as many screws as possible in order to reduce losses at the joints. Bandpass tuning is accomplished by varying the capacitance of the resonators with the five adjustable screws opposite each of the five fingers. This particular filter was designed with the aid of the low-pass prototype synthesis methods outlined in [4]. It is a 0.05-dB equal-ripple bandpass filter with an equal-ripple bandwidth of 0.15 GHz centered at 8.5 GHz.

Three different fabrication methods were considered: (1) making the entire filter out of bulk T_c superconductor; (2) coating a silver-plated copper filter with a thick YBCO film; and (3) coating a pure silver filter with a thick superconducting film. Despite experience with bulk high T_c structures [5], the first method, after some initial trials, was abandoned due to the complexity of the structure (many sharp edges, screw holes, threads, etc.). Some of the edges tended to chip, and the drilling of so many holes weakened or even destroyed the structure. In the second approach a copper filter was machined and silver-plated with a 500-μm thick film. This filter was subsequently coated with a thick YBCO film. The resultant film looked dark gray and was not superconducting. This process was repeated with a number of test pieces without the desired success. A possible explanation of this may be that the silver buffer layer degraded at high temperatures and the copper substrate reduced the thick YBCO film. The same result was obtained even when the sintering temperature was reduced from 920 deg C to 900 deg C. Finally, the third method attempted was very successful. It involved machining the filter out of 99.9 percent pure silver, and subsequently coating the filter with a thick YBCO film, as in the previous method. The resulting films looked black and exhibited sharp superconducting transitions with T_c ~92 K. By using this last method, three different filters were fabricated and tested, making appropriate improvements each time.

The YBCO compound was prepared via a solid state reaction by using yttrium oxide, barium carbonate, and copper oxide. Stochiometric amounts of the constituent materials were mixed and ballmilled in methanol for 16 hr. The slurry was dried and vacuum calcined [6] at 800 deg C for 4 hr in an oxygen pressure of 2.7×10^2 Pa. Thick films were fabricated by mixing the YBCO powder with an organic solvent, and a dispersant was added to improve the rheological properties. The suspension was applied to the silver filter substrate and dried at 80 deg C for about 2 hr. The film was then sintered at 920 deg C for 4 hr in an oxygen partial pressure (PO_2) of 1.1×10^4 Pa and annealed at 450 deg C for 16 hr in 1 atmosphere of oxygen. Sintering in low PO_2 enhances [7] the sintering kinetics of YBCO. In addition, the melting point of silver is slightly higher in reduced PO_2. The resulting coating thickness was on the order of 50 μm.

An 8510B Hewlett Packard Network Analyzer with an S-parameter test set was used to precisely tune the filter and perform both the return and insertion loss measurements. The filter was originally tuned at room temperature to better than 20 dB of return loss across the frequency band of interest and subsequently cooled down to 12 K by using a closed-cycle refrigerator (CCR). Its temperature was monitored by two separate sensors attached to its body and the temperature-dependent data were taken as the filter was allowed to warm up slowly. The insertion loss of the coaxial lines inside the CCR are substrated from the data presented. These lines were separately characterized as a function of temperature for the frequency band of interest.

For an equal-ripple Chebyshev filter, the insertion loss (L_s) at midband is given by the expression:

$$L_s \text{ (dB)} = 8.686(C_n/WQ_u) \tag{1}$$

where Q_u is the unloaded resonator Q, W is the fractional bandwidth, and C_n is a coefficient determined by the filter order and its band ripple [2].

From Eq. (1), since $Q_u \propto 1/R_s$, it is readily seen that the insertion loss of a filter is proportional to the surface resistance of the material it is made of. For conventional superconductors like Pb (T_c = 7.2 K) and NbTi (T_c = 9.8 K) at 4.2 K and at X-band frequencies, the surface resistance is known to be as much as three orders of magnitude lower than that of copper. The expected insertion loss for an ideal filter like the one considered here is therefore on the order of 10^{-4} dB. From the results on the surface resistance of polycrystalline and single crystal YBCO

materials [1], the authors expect insertion losses on the order of 0.25 dB for the polycrystalline and no more than 10^{-3} dB for an ideal single-crystal filter.

Figure 2 shows the temperature dependence of the insertion loss of the three different silver and YBCO filters as a function of temperature. All the data were taken at an input continuous wave (cw) power level of 38 μW. Qualitatively, all the curves look similar. A sharp transition at 93 K is observed as the film becomes superconducting. The insertion loss then tails off at about 70 K, at which point it starts decreasing linearly to 12 K. The lowest value reached is 0.69 dB, and if extrapolated down to 4.2 K, the insertion loss would be 0.61 dB, as compared with 0.55 dB for a similar copper filter. At lower values of input power, the filter exhibits losses lower than that of copper, as discussed later.

The data of Fig. 2 for the three filters that were constructed reveal an important feature, namely that the differences in performance among the three trials were not due to material properties (i.e., R_s), but rather due to nonoptimization of the final devices. This is evident if one superposes the three curves by subtracting constant (temperature-independent) offsets, whence the three curves become identical. Thus, in practice, the insertion loss is represented by $L_m(T) = L_s(T) + L_0$, where L_0 is temperature-independent and arises from connector mismatch, tuning, etc. When this was realized, it was possible to achieve the best results with filter 3 by improving the ground-plane contacts, and by carefully assembling and tuning the filter at room temperature to the lowest insertion loss achievable.

Earlier work [3] with NbTi (T_c = 9.8 K) filters substantiates the above conclusions for the high T_c filter. In the NbTi filter, it was discovered that poor ground-plane contacts can contribute as much as 0.5 dB to the insertion loss at cryogenic temperatures and could be minimized by using knife edges at the joints. The ultimate residual loss achieved with the NbTi filter was 0.10 dB, and it was concluded that this represented the loss due to the connectors.

Thermal cycling strongly affects the L_s data. For filter 3, L_s was found to increase after the first thermal cycle and following the filter assembly. After the first thermal cycle, L_s increased by 0.7 dB at room temperature, and the subsequent cool-down data showed an increase by the same amount. Following the final cryogenic measurement, it was found that room temperature disassembly and reassembly of the filter increased its loss by 3.0 dB. Thus, for future designs, care must be taken to reduce warping and dimensional changes in the structure due to the high temperatures involved in the fabrication process. In addition, engineering design changes will have to be incorporated to eliminate the contact losses at the ground planes.

The surface resistance of samples prepared in the exact same way as the filter was also measured as a function of temperature down to 77 K by using a 14-GHz microwave cavity, with the sample disk as an end plate. The R_s data showed the same qualitative temperature dependence that was observed for the insertion loss measurements.

As mentioned earlier, the insertion loss L_s was found to exhibit a strong input power dependence, even at very low input powers. Figure 3 shows L_s as a function of input power which was varied from 38 μW to 40 pW. For the first cool down and very low input cw power levels ($>$ 1 nW) the filter loss approached very low values of about 0.05 dB. As indicated in Fig. 3, the maximum error at the low power levels and for very low values of L_s is about 0.3 dB, which represents a very conservative upper bound on the measurement.

Such strong dependence on the input power is somewhat puzzling, since the incident magnetic fields on the superconducting film are too weak to account for such an effect. A thermal gradient could, however, exist in the YBCO material at the ground-plane contacts. Although the silver surface was machined flat to within 0.001 in., irregularities caused by firing and uneven film thickness prevented perfect mechanical contact at the interface, and hence led to poor electrical and thermal performance. It is likely that power levels above 38 nW could be sufficient to locally heat the YBCO material at the shorted base of the resonators a few kelvin above the rest of the material. It should be recalled that the radio frequency current densities are highest at the shorted base of the resonators. This is consistent also with the observed thermal cycling degradation, since the stainless steel screws would have a slightly different coefficient of expansion than the silver body, leading to weaker contact force and hence higher losses.

Note several features of the device reported here that bear on comparisons with thin-film microstrip filters. As part of its design, the waveguide should possess the lowest losses achievable, in comparison with microstrip filter structures, which have higher losses. The five element configuration also provides a high out-of-band rejection (~80 dB), which is difficult to achieve with thin-film microstrip filters. Connectorization should, in principle, be easier here because of the metallic substrate. It was also noted that the very low insertion loss of the filter has forced careful evaluation of the procedures for measuring L_s, and indeed it is clear that even better performance will require more exact procedures. It should also be noted that

at least in space communications, signal power levels are usually very low ($\sim$ nanowatts), which is the level at which this filter exhibits its lowest insertion loss (see Fig. 3).

The authors have also tested the filter in an actual system configuration designed to measure system noise temperature with a HEMT front-end amplifier. Details of the measurement will be presented elsewhere. At a physical system temperature of $T_m = 20$ K, and with the amplifier noise temperature of $T_L = 20$ K, the experiments yielded a noise temperature contribution of 4.6 K, which agrees well with that inferred from the measured insertion loss data.

In conclusion, a stripline high T_c superconducting microwave bandpass filter was successfully fabricated and tested. The lowest insertion loss measured was 0.05 dB, with an input return loss of better than 16 dB across the passband. The filter provides significant improvements over a comparable copper filter and is at present limited not by the superconducting material, but rather by design limitations possibly originating at the time of fabrication. Even at present, the filter is comparable to low T_c superconducting devices (with a distinct advantage over the latter in that operation is possible at elevated temperatures which are cost effective), and meets design criteria for the very low-noise communication systems of deep space applications.

Work at Northeastern University was supported by the Caltech President's Fund and by the Northeastern Center for Electromagnetics Research.

References

[1] D. H. Wu, L. L. Kennedy, C. Zahopoulos, and S. Sridhar, "Characteristics and Growth of Single Crystals of $Y_1Ba_2Cu_3O_7$ with Superior Microwave Properties," *Appl. Phys. Letters*, vol. 55, no. 7, pp. 696–698, August 14, 1989.

[2] A. Inam, et al., "Microwave Properties of Highly Oriented $YBa_2Cu_3O_{7-x}$ Thin Films," *Appl. Phys. Letters*, vol. 56, no. 12, pp. 1178–1180, March 19, 1990.

[3] J. J. Bautista and S. M. Petty, "Superconducting NbTi and Pb (Cu) Bandpass Filters," *IEEE Trans. on Magnetics, MAG-21*, no. 2, pp. 640–643, March 1985.

[4] G. L. Matthaei, L. Yound, and E. M. T. Jones, *Microwave Filters, Impedance-Matching Networks, and Coupling Structures*, Dedham, Massachusetts: Artech House Books, pp. 421–427, 1980.

[5] C. Zahopoulos, W. L. Kennedy, and S. Sridhar, "Performance of a Fully Superconducting Microwave Cavity Made of the High T_c Superconductor $Y_1Ba_2Cu_3O_x$," *Appl. Phys. Letters*, vol. 52, no. 25, pp. 2168, June 20, 1988.

[6] U. Balachandran, R. B. Poeppel, J. E. Emerson, S. A. Johnson, M. T. Lanagan, C. A. Youngdahl, D. Shi, and K. C. Goretta, "Synthesis of Phase-Pure Ortharhombic $YBa_2Cu_3O_x$ Under Two Oxygen Pressure," *Materials Letters*, vol. 8, no. 11, 12, pp. 454–456, November 1989.

[7] N. Chen, D. Shi, and K. C. Goretta, "Influence of Oxygen Concentration on Processing $YBa_2Cu_3O_{7-x}$," *J. Appl. Phys.*, vol. 66, no. 6, pp. 2485–2488, September 5, 1989.

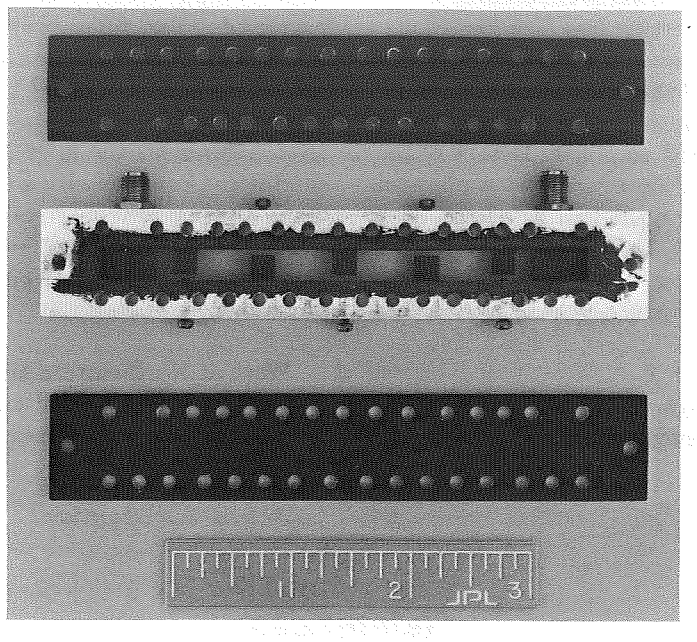


Fig. 1. The superconducting YBCO/Ag filter.

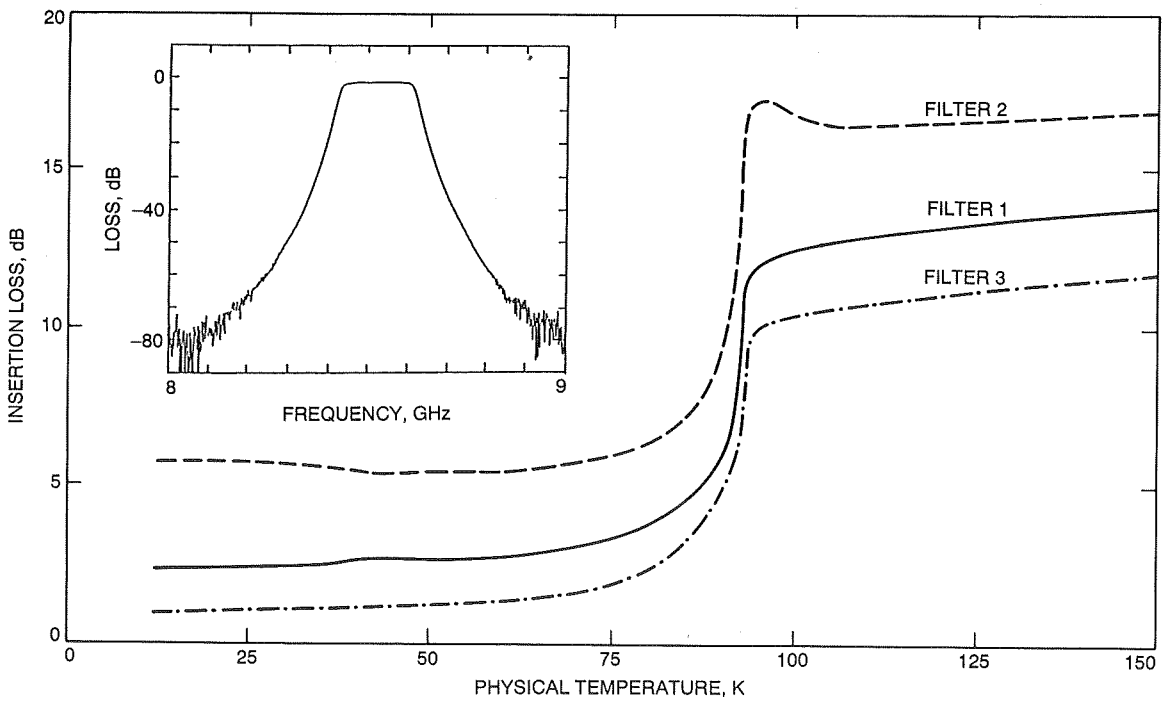


Fig. 2. Insertion loss (L_S) versus physical temperature (T) for three separate filters. The measurements were taken at an input power of 38 μW.

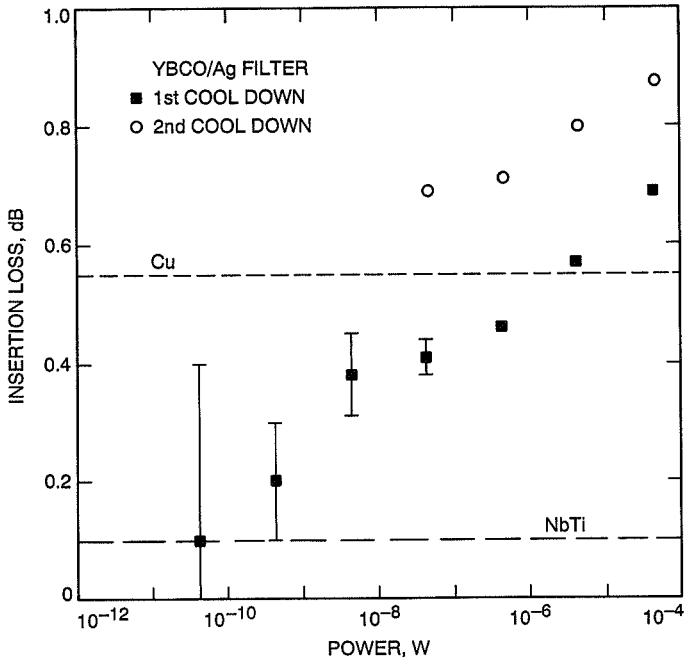


Fig. 3. Insertion loss measured at various power levels at 12 K for the YBCO/Ag filter, and similar Cu and NbTi filters.

Equivalent Circuit Model of Traveling-Wave Maser Slow-Wave Structures

J. Shell
Radio Frequency and Microwave Subsystems Section

An approach is presented for deriving transmission line equivalent circuits that can approximately model the S-*parameter response of traveling-wave maser slow-wave structures. The technique is illustrated by computing the* S-*parameter responses of an X-band and S-band maser slow-wave structure and comparing these with experimental measurements.*

I. Introduction

Since the early days of maser development, the traveling-wave maser (TWM) has been widely used because of its small size and its potential for relatively large instantaneous bandwidth, as compared with cavity masers. Several different slow-wave structure (SWS) geometries have been used, including meander lines, quarter-wavelength resonant lines, and half-wavelength resonant lines. One feature they all have in common is the use of predominantly transverse electromagnetic (TEM) resonators placed parallel to one another and coupled by their mutual capacitance and inductance.

Further developments in this area, including new resonant conductor geometries and impedance-matching networks that transition between the SWS and conventional transmission lines, would be enhanced by the use of an approximate equivalent circuit model for the SWS. The purpose of this article is to present one approach to arriving at such approximate equivalent circuits. Equivalent circuits provide a nice physical picture of the SWS and describe well its "filter" behavior: such properties as the upper and lower cutoff frequencies and the group delay. The computational time required to analyze a given structure is generally on the order of minutes, rather than hours or days, as required by many numerical procedures that compute the electromagnetic fields. Predicting the filter properties from an equivalent circuit is a logical first step in the design process. Ultimately, the electromagnetic fields must be calculated to ascertain other properties of the maser, such as the polarization of the signal and pump radio frequency (RF) magnetic fields inside the structure. However, this is best done when the choice of SWS geometries has already been narrowed from consideration of their filter behavior.

This article is organized into eight sections. Following the Introduction, Section II discusses what is required to predict the behavior of coupled-line circuits. Section III describes the SWS geometries that one can consider and some assumptions used in deriving an equivalent circuit. Section IV discusses the solution of the electrostatic field problem, which is necessary to calculate the capacitance matrix. Section V discusses the estimation of the effective dielectric constant, line lengths, and phase velocity. Section VI discusses one way of estimating the terminat-

ing line capacitances. Section VII presents two examples of the analysis: the current JPL X-band (8.45 GHz) half-wavelength comb SWS (Block IIA) and an S-band (2.3 GHz) quarter-wavelength comb SWS partially loaded with ruby. Part VIII presents some discussion and conclusions. Appendix A presents a few other equivalent circuits for representing coupled lines. Appendix B contains a listing of the TOUCHSTONE programs used for the examples in Section VII.

II. Coupled-Line Circuit Theory

Distributed elements, such as transmission lines, obey equations similar to ordinary reactances under certain circumstances. This requires the introduction of a new frequency variable $S = j\tan(\pi f/(2f_0))$. If f_0 is the frequency at which the line is a quarter-wavelength long, then $\pi f/(2f_0)$ is the electrical length of the line.

Consider the input impedance at one end of a transmission line, with the other end short-circuited. The input impedance is given by

$$Z_{in} = jZ_0 \tan\theta \tag{1}$$

where θ is the electrical length of the line. Rewritten in terms of the variable S, this equation is simply

$$Z_{in} = SZ_0 \tag{2}$$

The analogy with the impedance of a conventional inductor, $Z = j\omega L$, leads one to associate L with Z_0, and S with $j\omega$.

If the other end of the line is terminated in an open circuit, the input impedance is given by

$$Z_{in} = -jZ_0 \cot\theta \tag{3}$$

Rewritten in terms of the variable S, this equation is simply

$$Z_{in} = Z_0/S \tag{4}$$

The analogy with the impedance of a conventional capacitor, $Z = 1/(j\omega C)$ leads one to associate C with $1/Z_0$, and S with $j\omega$.

The impedance of a TEM transmission line is uniquely given by its static capacitance/unit length as

$$Z_0 = 1/(V_p C_{\text{static}}) = \frac{377 \text{ ohms}}{\sqrt{\epsilon_r} C_{\text{static}}/\epsilon_0} \tag{5}$$

where V_p is the phase velocity, 377 ohms is the impedance of free space, ϵ_r is the relative dielectric constant, and ϵ_0 is the permittivity of free space. The static capacitance network provides the connection between the field solution and the conductor geometry. It is not an equivalent circuit. The equivalent circuit is obtained by applying the relevant boundary conditions to the admittance or impedance matrix description of the coupled lines. This results in a set of equations that relate the static capacitance values to the L's, C's, and unit elements of the equivalent circuit. Again, the L's are a shorthand notation for shorted transmission lines and the C's are a shorthand notation for open-circuited lengths of transmission line.

When the problem is generalized to coupled lines, the voltage and current become two component vectors. They are related by

$$\frac{dV}{dz} = -j\omega LI \qquad \frac{dI}{dz} = -j\omega CV \tag{6}$$

where L and C are 2-by-2 matrices. C is given by

$$C = \begin{bmatrix} C_1 + C_m & -C_m \\ -C_m & C_2 + C_m \end{bmatrix} \tag{7}$$

Here, C_m is the mutual capacitance/unit length between the lines, and C_1 is the self capacitance/unit length between line 1 and the ground. Since a TEM line has a unique propagation velocity, the matrix product LC is diagonal, where the diagonal elements are just the square of the phase velocity. Therefore, the capacitances and the phase velocity are all one needs to know to describe the coupled lines.

Since the distributed capacitance/unit length between a line and its ground plane (and the phase velocity) can describe a TEM transmission line, it seems reasonable that the distributed mutual capacitance/unit length between two lines might also describe a transmission line. This approach has been developed and yields the three-line equivalent circuit for coupled TEM lines, shown in Fig. 1, as described by Seviora [1].

This article uses a different equivalent circuit to represent the coupled lines. (A few more alternative equivalent circuits are given in Appendix A.) The equivalent circuit chosen for this article was originally developed by Sato and Cristal [2]. (See Fig. 2; the bottom portion of the figure is the shorthand notation used in this article. Shorted lengths of transmission line are written as inductors, and the ground planes or return conductors are omitted.) The reader may notice that the transmission lines running diagonally have negative impedance. The S parameters of a negative characteristic impedance line are the complex conjugate of those of the positive impedance line. As long as the network analysis program correctly accounts for this, such negative impedances are not a problem.

In this article, all the "inductors" are shorted lengths of transmission line. Some of the "capacitors" here are capacitances/unit length and represent a distributed coupling that is modeled as a transmission line. However, the fringing capacitances discussed here are of the ordinary, lumped variety, and this mixture of capacitor types should be kept in mind by the reader.

To summarize, the static capacitance matrix is central to the theoretical description of microwave filters using parallel coupled lines. It serves a dual purpose. From the microwave viewpoint, the capacitance represents the distributed coupling between two lines, or between a line and a ground plane. A proper network description of this coupling requires familiarity with Richard's transformation and the Kuroda transformations. This article will not derive the network-equivalent circuit but will take it as given. From the physical viewpoint, the capacitances represent the usual capacitance found by solving the electrostatic field problem. The physical basis of the dual function of the static capacitance is a somewhat subtle idea, which involves the use of complex potentials [3], and will not be explored further here.

The original derivation of the impedance matrix for coupled lines was done by Jones and Bolljahn [4]. Riblet [3] put the admittance description on firm theoretical ground. An important set of papers by Wenzel [5–8] provides a good introduction to capacitance matrices and coupled-line theory.

III. Possible SWS Transverse Geometries and Assumptions

The range of SWS transverse (the cross-sectional view normal to the lengths of the lines) geometries that can be analyzed is large, if the program which calculates the capacitance matrix is of sufficient generality. The recent availability of software (hereafter denoted by MPMCTL) developed by workers at Syracuse University [9] enables a wide range of geometries to be considered. MPMCTL is written explicitly to handle the transverse geometries shown in Fig. 3. These include (a) infinitely thin coupled microstrips, (b) finite thickness coupled microstrips, (c) coupled suspended substrate striplines, (d) coupled rectangular conductors, (e) coupled striplines, (f) broadside coupled lines, and (g) a coplanar waveguide. However, user-defined geometries are also available. Multiple dielectric layers are allowed if the dielectric interfaces are parallel to the ground planes.

Since the analytic description is based on a two-dimensional static capacitance matrix, it implicitly assumes TEM-mode propagation. If the conductors are parallel to the ground planes and embedded in a material with a uniform dielectric constant, then it is a good assumption that the waves excited on the lines are TEM waves. If the dielectric loading is not homogeneous, the fields are not pure TEM. However, in that case, this article will calculate an effective homogeneous dielectric constant and continue to assume TEM-mode propagation.

Besides a homogeneous dielectric constant, the largest cross-sectional dimension of the line must be small, as compared with a wavelength, so that the axial components will be small, as compared with the transverse components, and the wave essentially TEM. If the lines are close together, as compared with a wavelength, one may neglect retardation effects as well.

The approach taken in the analysis is the following. The impedances of all the transmission lines in the equivalent circuit are determined from the static capacitance matrix (calculated using MPMCTL) for that line geometry. The electrical length of the lines is determined from the known physical length and the effective dielectric constant. The terminating reactances are calculated separately. For fringing capacitances, the author uses the published results of Getsinger [10]. The S-parameters of the final equivalent circuit are calculated with a microwave computer-aided design (CAD) program called TOUCHSTONE [11].

IV. Solution of the Electrostatic Field Problem

The static capacitance matrix discussed in Section III is obtained by solving the two-dimensional Laplace equation for the static electric field for the given arrangement of conductors (assuming that the conductors are infinite

in length). There are several approaches to solving such boundary value problems. The approaches will not be discussed in detail, because they are covered extensively in the literature. Nevertheless, a few remarks are in order.

Conformal mapping has been used to obtain exact results when there are two conductors symmetrically placed between ground planes [12]. If they are unsymmetrically placed, the analysis becomes very unwieldy. The numerical approaches are able to handle a larger number of conductors than the analytical approaches. This is especially important if next-nearest-neighbor coupling is significant. The finite difference-equation approach has been extensively used [13], but it requires more computer time than other techniques and, unless a very fine mesh is used, is generally not extremely accurate. Its primary advantage is the simplicity of the algorithm.

The most accurate results, especially for edge-coupled striplines, have been obtained by using a Green's function integral equation moment method. This moment method can give accuracies of a few tenths of a percent [14]. The calculation of the capacitance matrix using such a method was carried out by Kammler [15]. This approach is general enough to handle any number of conductors; however, it is restricted to very thin conductors that are parallel to the ground planes and embedded in a homogeneous dielectric medium. The current JPL X-band masers (Block IIA) satisfy these restrictions quite well. A FORTRAN program based on Kammler's algorithm was written and used by the author to model the X-band maser prior to the appearance of MPMCTL.

The MPMCTL software uses a free-space Green's function integral equation formalism [16]. The dielectric is replaced by the equivalent bound charge densities. The equations are solved by using a moment method that gives good agreement with the author's program in the geometries common to both. MPMCTL also gives the charge densities on each pulse. MPMCTL is general enough to handle losses, although that capability is not used for this article. Further improvements to the model will include this.

V. Effective Dielectric Constant, Relative Velocity, and Line Length

The effective relative dielectric constant can be regarded as the ratio of the charges on a given conductor with partial dielectric loading to the ratio of the charges on the same conductor when there is no dielectric, provided that the potentials of the conductor are the same in both cases. Since the capacitance is the charge per unit voltage, if one assumes always a voltage of 1 V, the elements of the capacitance matrix give the charges on the conductors. That is, C_{ij} is the charge on conductor i if conductor j is at a potential of 1 V and all other conductors are grounded.

It is particularly easy to determine the effective dielectric constant at the lower cutoff frequency, the midband frequency, and the upper cutoff frequency, since the voltage distribution on the conductors is simple. At the lower cutoff frequency, all the conductors are at the same potential (for example, 1 V). At midband, the voltage progresses from finger to finger as +1 V, 0 V, −1 V, 0 V, +1 V, etc. At the upper cutoff frequency the voltage progresses as +1V, −1 V, +1 V, −1 V, etc. (See Fig. 4.)

Consider the following example. If one has three conductors, as shown in Fig. 4(a), then the charge on conductor 2 (i.e., Q_2) equals C_{21} plus C_{22} plus C_{23}. Note that C_{21} and C_{23} are negative, that is, a 1-V potential on fingers 1 and 3 induces a negative charge on finger 2. A 1-V potential on finger 2 induces a positive charge on finger 2. The total charge on finger 2 is the algebraic sum of the charges.

If the three conductors are at midband, Fig. 4(b), then the charge on finger 2 is given by C_{22} alone, because fingers 1 and 3 at 0 V do not induce any charge on finger 2. Therefore, Q_2 equals C_{22}.

If the three conductors are at the upper cutoff frequency, Fig. 4(c), then the negative potential of fingers 1 and 3 induces a positive charge on finger 2, which adds to the positive charge induced by finger 2 being at 1 V. Therefore, Q_2 equals $|C_{21}|$ plus C_{22} plus $|C_{23}|$.

This is repeated for the partially loaded geometry, and the effective relative dielectric constant is given by the ratio of Q_2 in the loaded and unloaded cases. That is, ϵ_{eff} equals $Q_2(\text{loaded})/Q_2(\text{empty})$.

Once the effective dielectric constant is known, the phase velocity is equal to $c/\sqrt{\epsilon_{\text{eff}}}$, where c is the speed of light in a vacuum. An estimate for the electrical length of the fingers in degrees can now be found from

$$\theta = (360 f l_p)/V_p \tag{8}$$

where f is the frequency, l_p is the physical length, and V_p is the phase velocity in the medium of propagation. Determining this line length is necessary for the equivalent

circuits. The physical length is known from the geometry of the structure.

VI. Terminating Reactances

The masers built at JPL generally use a capacitive loading on one or both ends of the fingers. The capacitances used in practice are very small, on the order of tenths of picofarads. Although these capacitances are small, they play a vital role. One might guess that a finger a full half-wavelength long, with no terminating capacitance, could also serve as a possible SWS geometry. However, in this case, the capacitive coupling between the lines exactly balances the magnetic coupling, and there is no propagation down the SWS. The capacitive loading at the ends of the fingers is necessary to upset the balance.

Another way to see this is to examine the Poynting vector. The transverse E and H fields account for the propagation of energy back and forth along the finger. They do not lead to energy propagation from finger to finger down the SWS. The component of the fringing electric field along the length of the finger (see Fig. 5) and the magnetic field perpendicular to the length of the finger cause a component of the Poynting vector to be directed down the structure.

To model SWS's, one must estimate the capacitive loading from the fingertips to the surrounding copper walls. This capacitance is generally some combination of parallel plate capacitance and fringing capacitance. (A fringing capacitance accounts for the electric field lines that are not straight and uniform between the conductors, but rather tend to bulge near the ends.) There are two fingertip loading capacitances that must be estimated; they are shown in Fig. 6. The first, designated C_f, is the capacitance from the end of the finger to the surrounding walls. The second, designated C_x, is the capacitance from the end of one finger to the end of an adjacent finger.

It should be mentioned that these capacitances will depend on the frequency within the passband of the SWS. For example, at the lower cutoff frequency, all the fingers are at the same potential. The electric field has no normal component at the plane midway between the fingers. Therefore, C_x must be zero at the lower cutoff frequency. At the upper cutoff frequency, adjacent fingers are at opposite potential, and the electric field has only a normal component at the plane midway between the fingers. Here C_x must be maximum. This article assumes that C_f and C_x are independent of the frequency. The purpose of this section is to show how to obtain ballpark estimates of the capacitances for use in the equivalent circuits. Further refinements would include more exact calculations, as well as the frequency dependance.

The values of the fringing capacitances can be estimated with the results obtained by Getsinger for rectangular conductors centered between ground planes. The capacitances that Getsinger calculates are shown in Fig. 7. The odd-mode fringing capacitance, C'_{f0}, is the capacitance per unit length to ground (this includes the top and bottom ground planes, as well as the end wall) from one corner and half the associated vertical wall in the coupling region for a bar with odd-mode excitation. The even-mode-fringing capacitance, C'_{fe}, is the capacitance per unit length to ground (in this case, only the top and bottom planes) from one corner and half the associated vertical wall in the coupling region of a bar with even-mode excitation. Getsinger determined the capacitances by using a conformal mapping technique. The results are exact for bars extending in width infinitely far from the coupling region between the conductors. (Future CAD work may benefit from expressions derived by Perlow [17], which provide relatively simple algebraic expressions for the coupling capacitances, in place of Getsinger's charts.)

In order to estimate C_f, use Getsinger's values of $C'_{f0}(s/b, t/b)$, where s is twice the gap (g) between the end of the finger and the opposing ground plane. See Fig. 8. Since the capacitances are per unit length, multiply by the width w of the conductor. Since Getsinger's charts are for centered conductors, estimate C_f by adding the capacitances for the two cases of different b. (The capacitances are added because each one gives the contribution from only half of the vertical wall at the end of the finger.) This gives the fringing capacitance from a horizontal edge as

$$C_f = \epsilon w \left\{ \frac{C'_{f0}}{\epsilon} (2g/b_1, t/b_1) + \frac{C'_{f0}}{\epsilon} (2g/b_2, t/b_2) \right\} \quad (9)$$

where ϵ, equal to $\epsilon_0\, \epsilon_r$, is the dielectric permittivity of the medium. Here ϵ_0, equal to 0.0886 pF/cm, is the permittivity of free space, and ϵ_r is the relative dielectric constant.

The finger-to-finger fringing capacitance C_x can be estimated from Getsinger's charts if one uses C'_{fe} and the geometry of Fig. 9. Now the spacing between the ground planes is the pitch p of the slow-wave structure. The width of the fingers w (as used above) now plays the role of Getsinger's t and vice versa. Thus, one obtains

$$C_x = \epsilon t \frac{C'_{fe}}{\epsilon} (2g/p, w/p) \quad (10)$$

There is no factor of 2 here because in the equivalent circuits one has two capacitors C_x leaving each fingertip. The values obtained in this way can be used as starting values for the models.

VII. Examples

A. X-Band Half-Wavelength Comb

The geometry of the current X-band Block IIA SWS is shown in Fig. 10. If one starts with the equivalent circuit of Sato and Cristal, this will model the coupling between the fingers. The nodes at the ends of the transmission lines in the equivalent circuit correspond to the ends of the fingers. The 12-mil gap between the ends of the fingers and the copper cavity (ground) is represented by a fringing capacitance C_f. The electric field between the ends of the fingers is represented by a fringing capacitance, C_x.

The capacitance matrix for an array of five conductors with the geometry of Fig. 10(a) can be calculated with MPMCTL and is given by

$$[C] = \begin{bmatrix} 272.2 & -43.04 & -3.996 & -0.876 & -0.265 \\ -43.04 & 280.8 & -42.27 & -3.83 & -0.877 \\ -3.99 & -42.27 & 280.8 & -42.27 & -3.99 \\ -0.876 & -3.83 & -42.27 & 280.8 & -43.04 \\ -0.266 & -0.877 & -3.99 & -43.04 & 272.2 \end{bmatrix} \text{pF/m} \tag{11}$$

The self and mutual capacitances of the parallel conductors are easily derived from the capacitance matrix. C_{self} is given by

$$\left.\begin{aligned} C_{\text{self}} &= -3.99 - 42.27 + 280.8 - 42.27 - 3.99 \\ &= 188.28 \text{ pF/m} \\ C_{\text{mut}} &= 42.27 \text{ pF/m} \end{aligned}\right\} \tag{12}$$

The line admittances are given by

$$Y_{\text{self}} = V_p C_{\text{self}} = \frac{(3 \times 10^8)(188.28 \times 10^{-12})}{\sqrt{9.8}}$$

$$= 1.80 \times 10^{-2} \text{ mhos} \tag{13}$$

or $Z_{\text{self}} = 1/Y_{\text{self}} = 55.5$ ohms

$$Y_{\text{mut}} = \frac{(3 \times 10^8)(42.27 \times 10^{-12})}{\sqrt{9.8}} = 4.05 \times 10^{-3} \text{ mhos} \tag{14}$$

or $Z_{\text{mut}} = 1/Y_{\text{mut}} = 247$ ohms.

The line lengths are easily determined because the dielectric filling is assumed to be uniform. For the X-band maser, the finger length of 0.196 in., see Fig. 10(b), corresponds to an electrical length of 150 deg at 8 GHz.

Finally, the fringing capacitances are determined. Using values of g equal to 0.012 in., s equal to 0.024 in., b_1 equal to 0.080 in., b_2 equal to 0.220 in., t equal to 0.002 in., ϵ_0 equal to 0.0886 pF/cm, w equal to 0.040, $\epsilon_r = 9.8$, and Getsinger's charts, one calculates C_f equal to 0.19 pF. By using g equal to 0.012 in., p equal to 0.080 in., t equal to 0.040 in., w equal to 0.002 in., ϵ_0 equal to 0.0886 pF/cm, ϵ_r equal to 9.8, and Getsinger's charts, one calculates C_x equal to 0.01 pF.

The equivalent circuit for five fingers is shown in Fig. 11. The author finds that better agreement with measured data is obtained if C_f is decreased slightly to 0.1 pF, and C_x is changed to -0.005 pF. A calculation of S_{21} for this circuit with the new values using TOUCHSTONE is given in Fig. 12. The measured behavior of S_{21} for a comb with 44 fingers is shown in Fig. 13. The predicted group delay for five fingers is shown in Fig. 14. At midband, the delay is about 1 nsec. Therefore, for 44 fingers one would expect about 8.8 nsec. The measured group delay is shown in Fig. 15 and is about 9 nsec at midband.

B. S-Band Quarter-Wavelength Comb

In this example, the basic approach used above can be generalized to handle more complicated structures. The example chosen is an S-band quarter-wavelength comb structure with ruby loading on one side of the fingers, and no dielectric on the other side. The dielectric does not extend along the entire length of the finger. The geometry is shown in Fig. 16. A similar geometry is being considered for a Ka-band (32-GHz) traveling-wave maser.

The problem naturally lends itself to a two-part analysis. The first part begins at the bottom of the cavity, where the fingers are shorted, and extends to the point on the fingers where the dielectric stops. The second part begins at the top of the dielectric and extends to the cover.

The first part can be thought of as a conventional comb structure, with the additional complication of partial dielectric loading. The first task is to derive an equivalent circuit for the comb structure by starting from the basic circuit of Sato and Cristal. Then the capacitance matrix is calculated as before, by using MPMCTL. The capacitance matrix will also be calculated for the empty comb, which is necessary in order to model the unloaded portion of the comb, as well as to calculate the effective dielectric constant. The line admittances will be derived from the elements of the first capacitance matrix as before, except that the velocity of propagation will depend on the value of the effective dielectric constant. The effective electrical length of the lines in degrees also depends on this effective dielectric constant. The second part of the circuit will be represented by a Sato/Cristal pair of lines. One set of nodes coincides with the nodes at the top of the dielectric, the other set of nodes is located at the tops of the fingers. As before, the electric field from the fingertips to ground and from fingertip to fingertip is represented by C_f and C_x, respectively.

Starting with the equivalent circuit for a pair of coupled lines, a somewhat simpler circuit results if one end of all the conductors is grounded. The resulting circuit is shown in Fig. 17. One is left with an equivalent circuit composed only of S-plane inductors (i.e., shorted transmission lines).

The capacitance matrix for the finger geometry without any dielectric is obtained by using MPMCTL. The result for three fingers is

$$[C] = \begin{bmatrix} 33.76 & -9.64 & -0.259 \\ -9.64 & 37.42 & -9.64 \\ -0.259 & -9.64 & 33.76 \end{bmatrix} \text{ pF/m} \qquad (15)$$

The capacitance matrix for the section of the structure with partial filling requires more work. The user-defined geometry portion of MPMCTL must be used. The 136 nodes used for the moment method calculation are shown in Fig. 18.

The capacitance matrix for three fingers is calculated to be

$$[C] = \begin{bmatrix} 107.6 & -24.95 & -0.800 \\ -25.02 & 117.1 & -25.00 \\ -0.800 & -24.98 & 107.6 \end{bmatrix} \text{ pF/m} \qquad (16)$$

Therefore, the effective dielectric constants at lower, midband, and upper cutoffs are given by

$$\left.\begin{aligned} \epsilon_{\text{eff}}(\text{lower cutoff}) &= \frac{67.08}{18.13} = 3.70 \\ \epsilon_{\text{eff}}(\text{midband}) &= \frac{117.1}{37.42} = 3.13 \\ \epsilon_{\text{eff}}(\text{upper cutoff}) &= \frac{167.12}{56.7} = 2.95 \end{aligned}\right\} \qquad (17)$$

The line impedances for the empty portion of the comb are determined as before,

$$\left.\begin{aligned} C_{\text{self}} &= -9.64 + 37.42 - 9.64 = 18.14 \text{ pF/m} \\ C_{\text{mut}} &= 9.64 \text{ pF/m} \end{aligned}\right\} \qquad (18)$$

$$\left.\begin{aligned} Y_{\text{self}} &= (3 \times 10^8)\,(18.14 \times 10^{-12}) \\ &= 5.44 \times 10^{-3} \text{ mhos} \\ Z_{\text{self}} &= 1/Y_{\text{self}} = 184 \text{ ohms} \end{aligned}\right\} \qquad (19)$$

$$\left.\begin{aligned} Y_{\text{mut}} &= (3 \times 10^8)\,(9.64 \times 10^{-12}) \\ &= 2.892 \times 10^{-3} \text{ mhos} \\ Z_{\text{mut}} &= 1/Y_{\text{mut}} = 346 \text{ ohms} \end{aligned}\right\} \qquad (20)$$

Similarly, for the partially filled portion of the comb,

$$\left.\begin{aligned} C_{\text{self}} &= -25.02 + 117.1 - 25.00 = 67.08 \text{ pF/m} \\ C_{\text{mut}} &= 25.0 \text{ pF/m} \end{aligned}\right\} \quad (21)$$

$$Y_{\text{self}} = \frac{(3 \times 10^8)\,(67.08 \times 10^{-12})}{\sqrt{3.13}} = 1.1375 \times 10^{-2} \text{ mhos} \quad (22)$$

$$Y_{\text{mut}} = \frac{(3 \times 10^8)\,(25.01 \times 10^{-12})}{\sqrt{3.13}} = 4.2409 \times 10^{-3} \text{ mhos} \quad (23)$$

$$Z_{\text{mut}} = 1/Y_{\text{mut}} = 236 \text{ ohms} \quad (24)$$

$$Y_{\text{self}} - Y_{\text{mut}} = 7.1341 \times 10^{-3} \text{ mhos} \quad (25)$$

$$1/(Y_{\text{self}} - Y_{\text{mut}}) = 140 \text{ ohms} \quad (26)$$

The line lengths are also required. For the unloaded portion of the line, θ equals 16.5 deg at 2 GHz. For the loaded portion, θ equals 47.5 deg.

The fringing capacitances are determined as before. Using $t = 0.050$ in., $s = 0.140$ in., $b = 0.280$ in., $w = 0.050$ in., one finds that $C_f = 0.0214$ pF. For the finger-to-finger fringing capacitance, one uses $t = 0.050$ in., $s = 0.140$ in., $b = 0.125$ in., and $w = 0.050$ in. One then finds that $C_x = 0.01$ pF.

The equivalent circuit for five fingers is shown in Fig. 19. Better agreement with the measured data is obtained if C_x is increased to 0.065 pF and the electrical length of the loaded portion of the fingers is increased to 56 deg. A TOUCHSTONE analysis of this circuit is shown in Fig. 20. The measured data are shown in Fig. 21. The calculated group delay for five fingers is shown in Fig. 22; at midband this is about 2 nsec. Therefore, 48 fingers would be expected to have a group delay of about 19.2 nsec. The measured data are shown in Fig. 23. The group delay is about 20 nsec at midband.

VIII. Discussion and Further Improvements

A summary of the the two examples worked out in Section VII is given in Table 1. The first three rows of the table compare the calculated and measured values of the upper and lower cutoff frequencies (at approximately the −3-dB points) and the group delay at the center of the bandpass. The calculated values are for a five-resonator structure. (The group delay is multiplied by the appropriate factor to make the number of resonators the same in the calculated and measured cases.) The values of the parameters used in the equivalent circuits are shown in the last eight rows of the table. If those values are different from those calculated in Section VII from purely geometrical considerations, the calculated value is given in parentheses next to the value used in the TOUCHSTONE program. Some flexibility was allowed because the theoretical filter response can be a very sensitive function of some parameters whose values are hard to calculate precisely. For example, the bandwidth is a very sensitive function of C_x; if one judged the model solely on the basis of the prediction by using the estimated value for C_x, one might conclude that the overall model is a poor one, when in fact the problem lies in the estimation of C_x.

The values for the line impedances were never adjusted, and only the electrical length for the partially filled portion of the S-band structure was adjusted. The value of 56 deg corresponds to a dielectric constant of about 4.3, which is not too different from the estimated effective dielectric constant of 3.7 at the lower cutoff. Nevertheless, this may be hinting that a more accurate model is really necessary, one which takes into account the quasi-TEM nature of the propagation.

Among the more interesting surprises was the need for a negative finger-to-finger fringing capacitance for the X-band model. The reason for this is still unclear. A positive value here can also give approximately the right bandwidth, but it was found that the negative value matches more nearly the observed overall shape of the group delay response. In particular, the greater slowing at the upper cutoff is predicted with a negative coupling capacitance, whereas a positive capacitance predicts a greater slowing at the lower cutoff.

Parallel coupled strip lines and rectangular conductors in a geometry useful for maser slow-wave structures can be approximately modeled. The models require calculation of the static capacitance matrix of the two-dimensional array of transmission lines, as well as any (reactive) terminations on the ends of the lines. From a knowledge of the capacitance matrix, transmission-line equivalent circuits for the coupled lines can be found. The (reactive) terminations are modeled as lumped elements. Approximate estimates of these terminations may be obtained by using published data, such as Getsinger's [10]. Once the circuit is obtained, the S-parameters can be obtained with a commercially available CAD program, such as TOUCH-

STONE. All the S-parameters can be calculated, as well as the group delay.

Further work in this area would include better calculation of the fringing capacitances, especially the variation of these capacitances with frequency For the inhomogeneous structures, a four-transmission-line equivalent circuit (as is shown in Appendix A) should give better results than the "equivalent dielectric constant" approach used in this article.

Acknowledgments

The author thanks R. J. Wenzel for bringing the work of Sato and Cristal to his attention, and to Professor Sam Derman of the City College of New York for numerous valuable discussions.

Table 1. Summary of X- and S-band SWS parameters

	X-Band		S-Band	
	Measured	Calculated	Measured	Calculated
f_{lower}, GHz	7.3	7.2	2.41	2.44
f_{upper}, GHz	8.4	8.54	2.86	2.90
Delay, nsec	9.0	8.8	20.0	19.2
	$Z_1(\Omega)$	55.5		346
	$Z_2(\Omega)$	247		184
	$Z_3(\Omega)$	—		236
	$Z_4(\Omega)$	—		140
	θ_1 (deg)	150		16.5
	θ_2 (deg)	—		56 (47.5)
	C_F(pF)	0.10 (0.19)		0.0214
	C_x(pF)	−0.005 (0.0099)		0.065 (0.01)

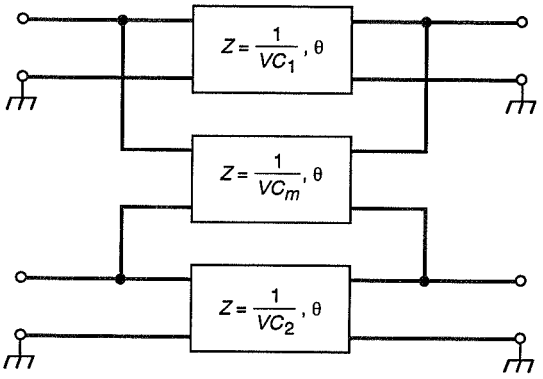


Fig. 1. Three-line equivalent circuit for coupled TEM lines.

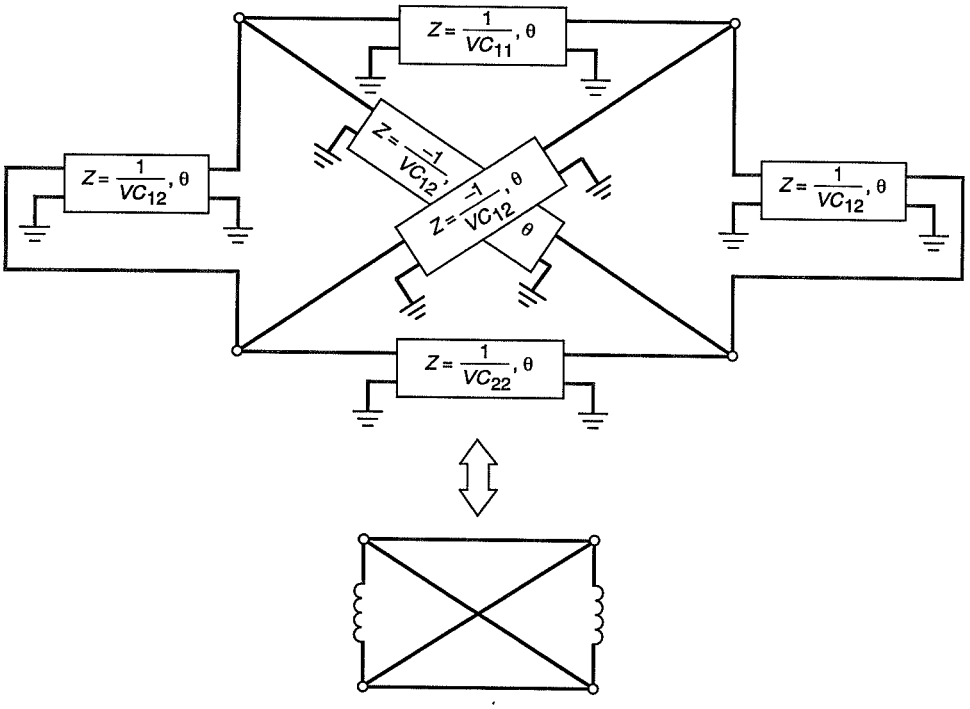


Fig. 2. Six-line equivalent circuit for coupled TEM lines.

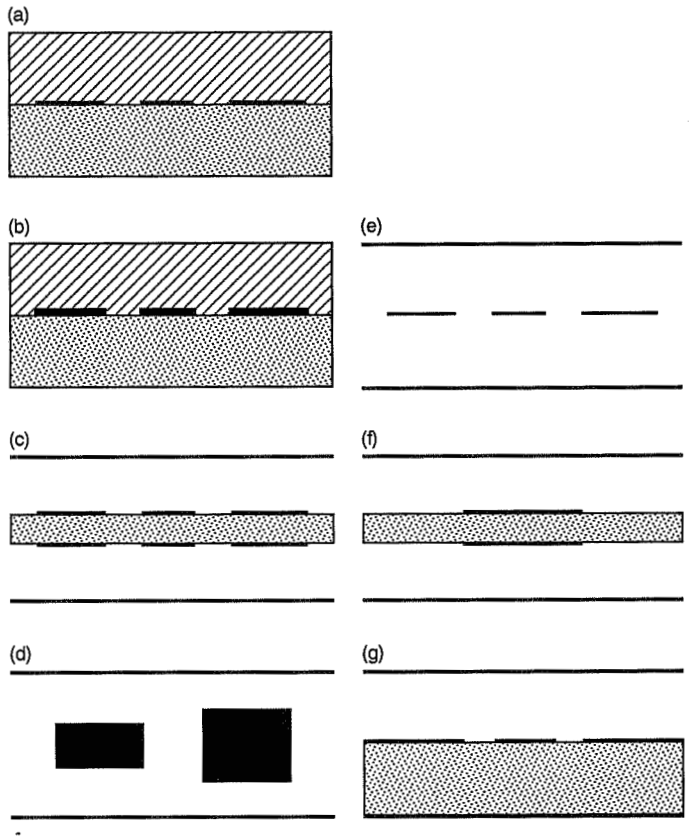


Fig. 3. Transverse geometries that can be handled by MPMCTL.

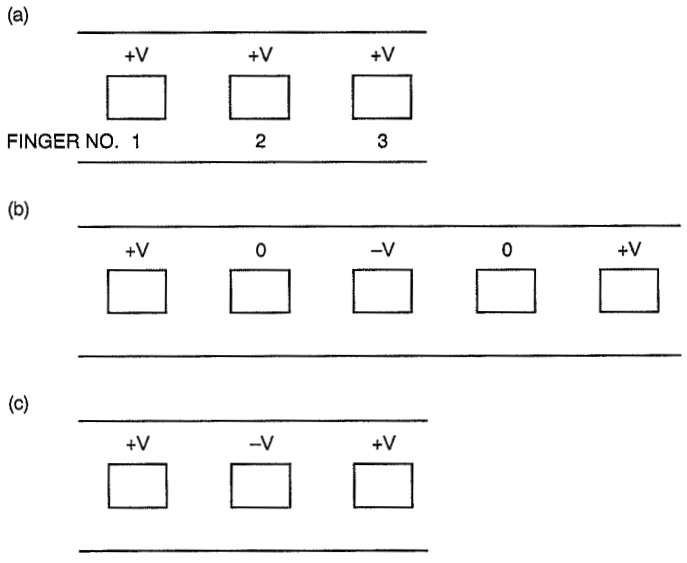


Fig. 4. Voltage distribution along the fingers for various circumstances: (a) at the lower cutoff frequency; (b) at the midband frequency; and (c) at the upper cutoff frequency.

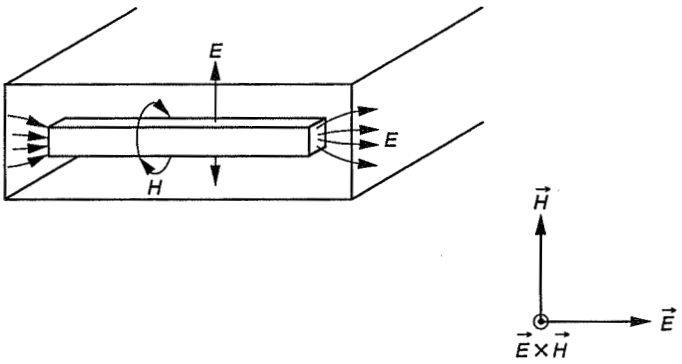


Fig. 5. Electric and magnetic field lines for a half-wavelength finger inside a cavity.

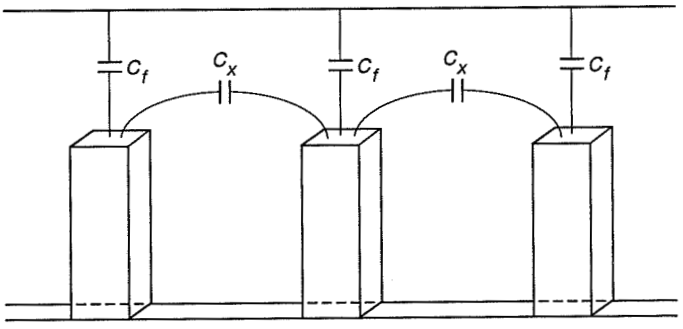


Fig. 6. The fringing capacitances from the ends of the fingers to ground (C_f) and between fingertips (C_x).

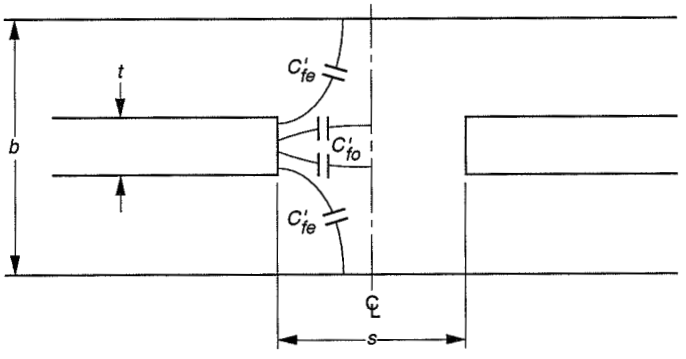


Fig. 7. The even-mode (C'_{fe}) and odd-mode (C'_{f0}) fringing capacitances considered by Getsinger.

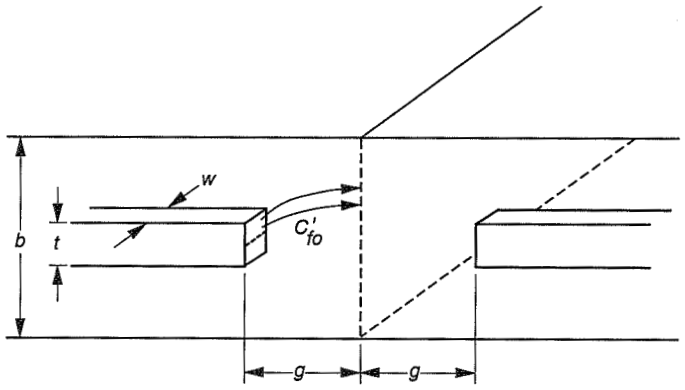


Fig. 8. Geometry used to calculate the fringing capacitance from the ends of the fingers to ground.

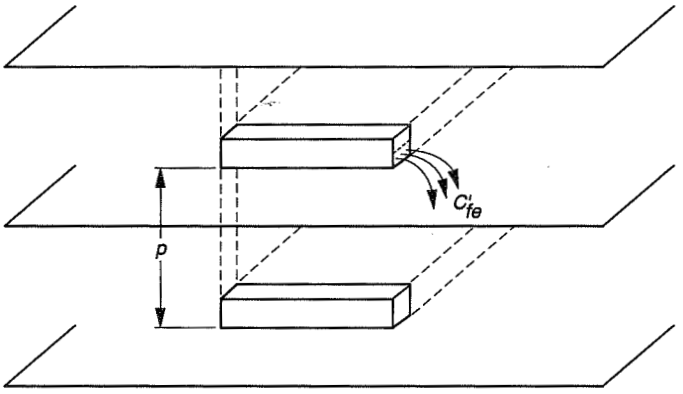


Fig. 9. Geometry used to calculate the fringing capacitance from fingertip to fingertip.

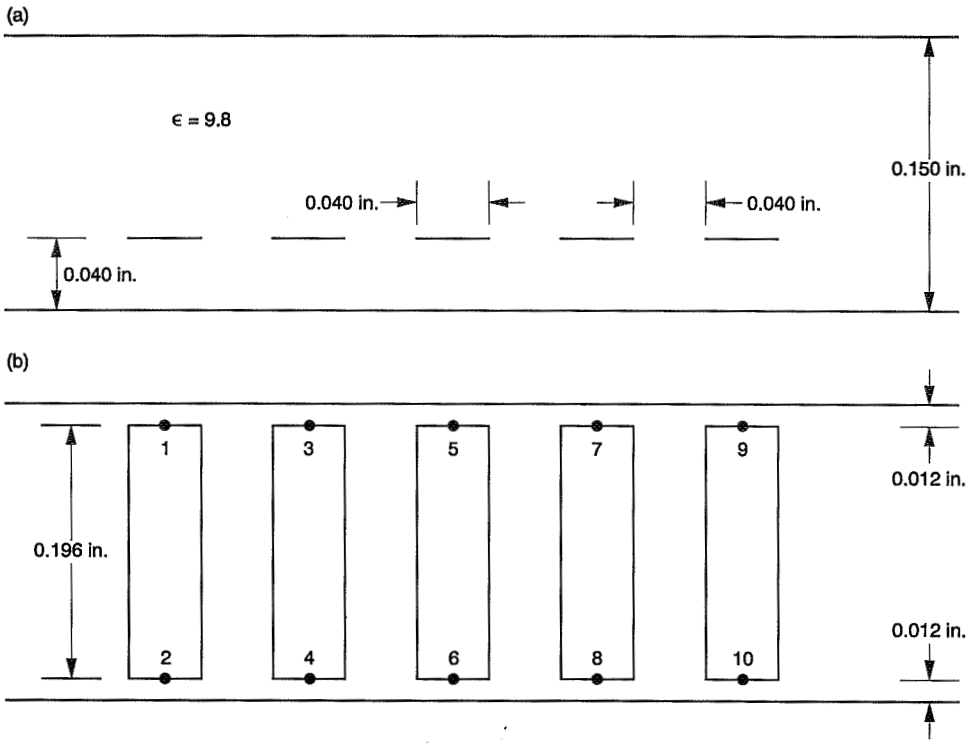


Fig. 10. Geometry of the X-band (Block IIA) slow-wave structure for five fingers: (a) transverse view, and (b) side view.

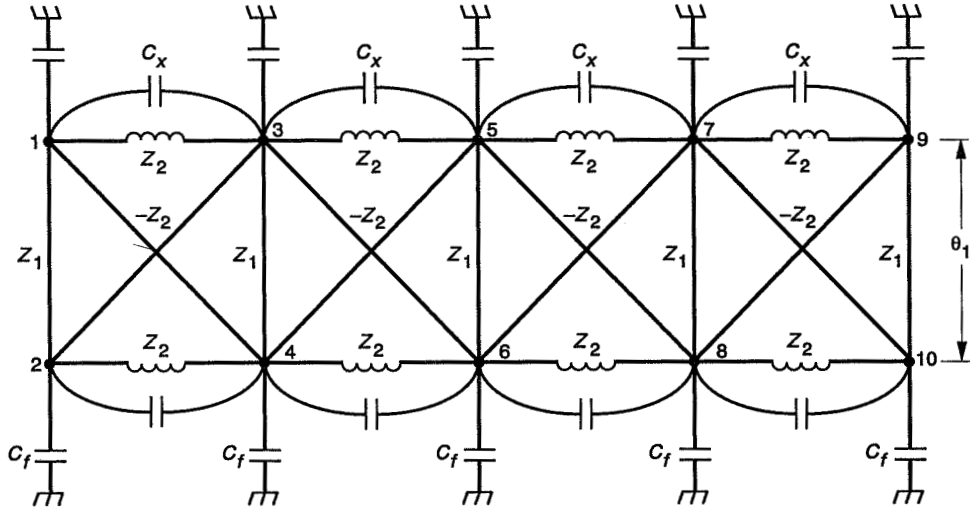


Fig. 11. Equivalent circuit used to model the X-band slow-wave structure of Fig. 10.

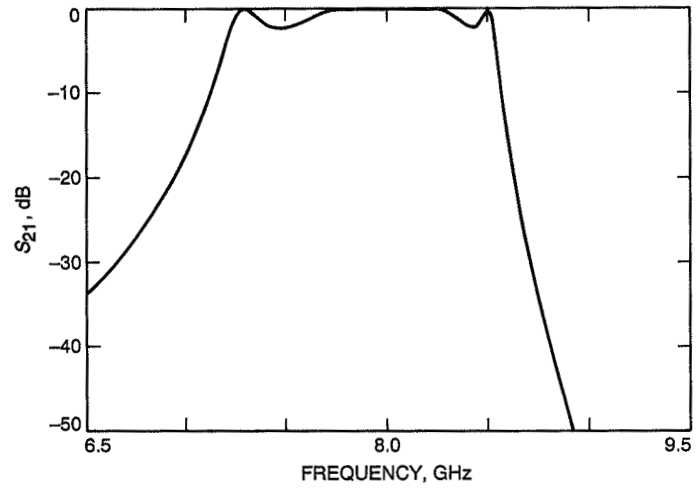


Fig. 12. Calculated S_{21} response for the circuit of Fig. 11.

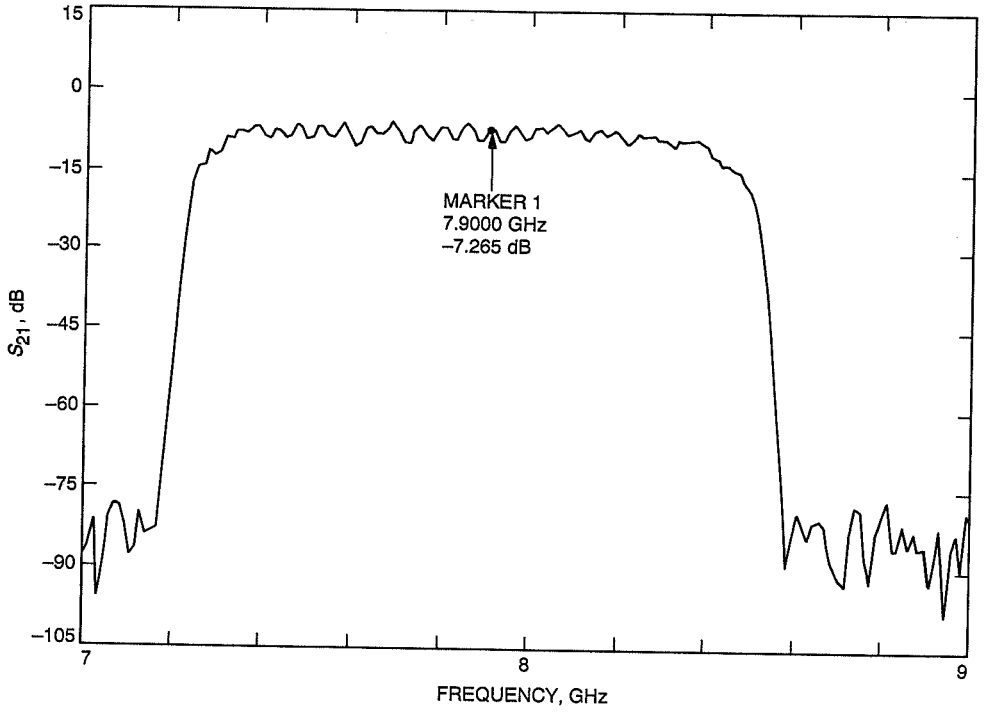


Fig. 13. Measured S_{21} response for the Block IIA slow-wave structure, which contains 44 fingers.

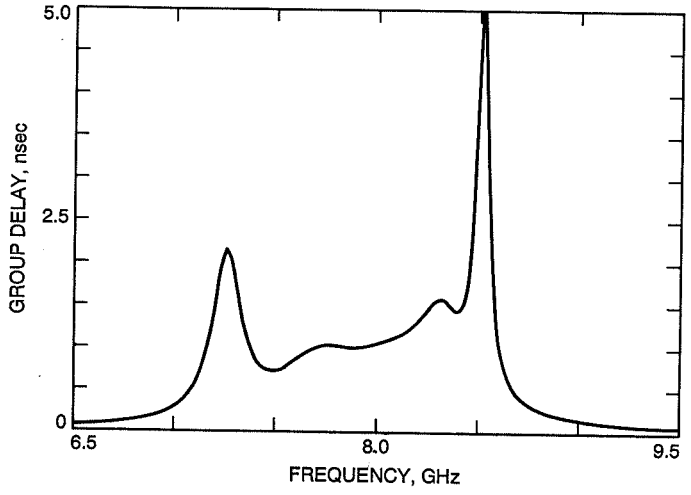


Fig. 14. Calculated group delay for the circuit of Fig. 11.

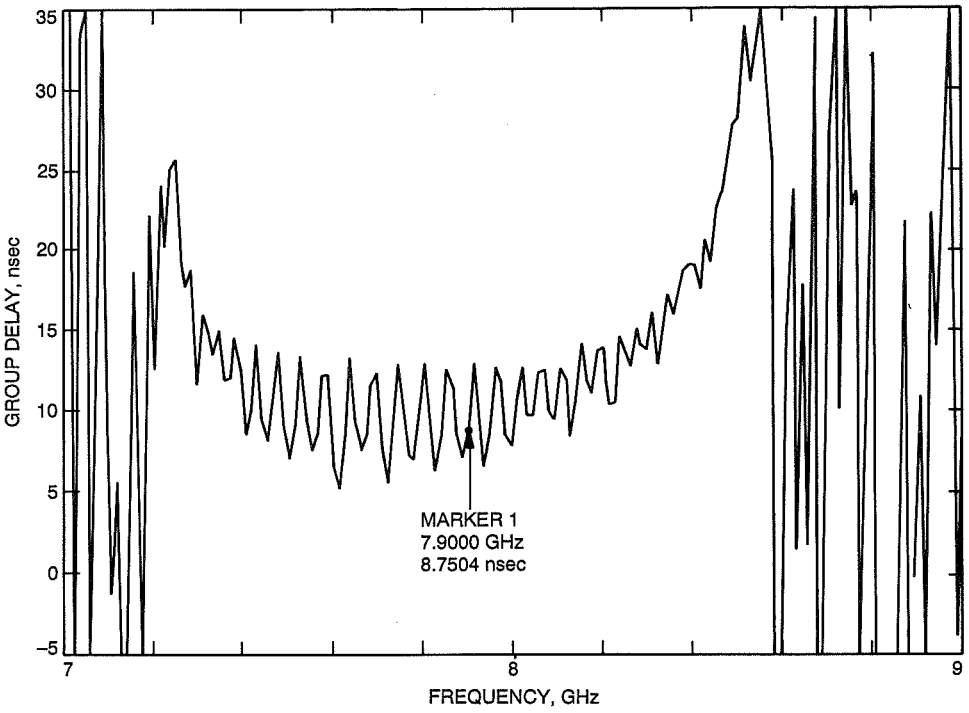


Fig. 15. Measured group delay for the Block IIA slow-wave structure.

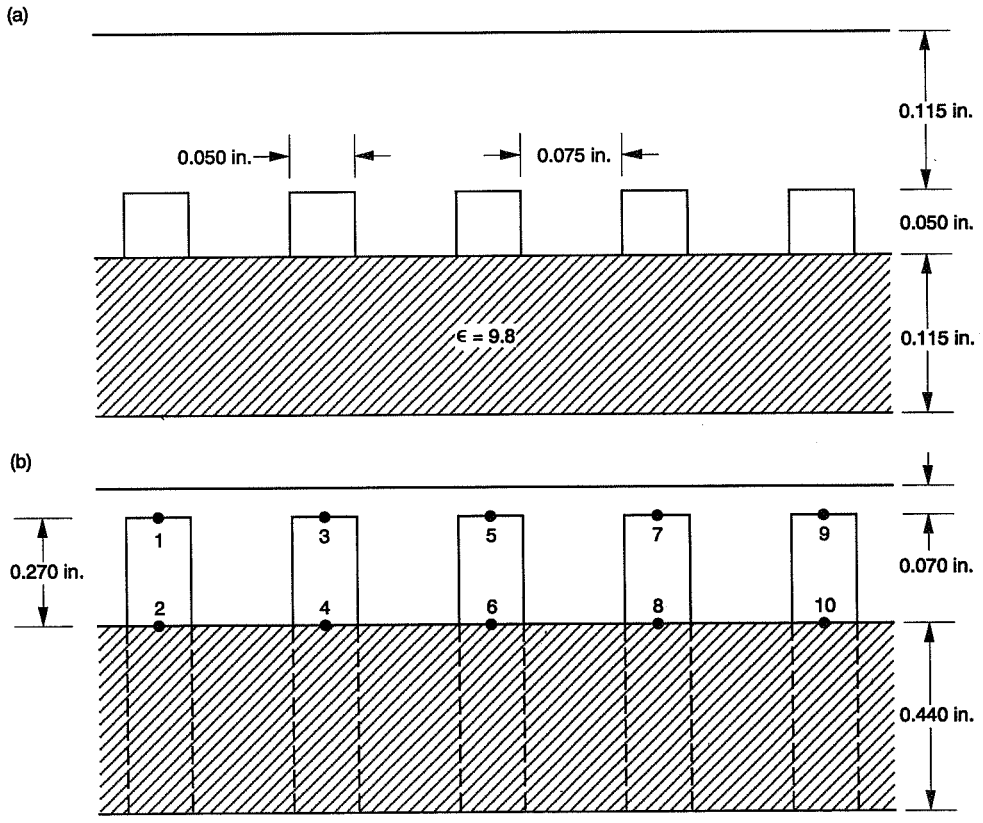


Fig. 16. Geometry of the partially loaded S-band slow-wave structure for five fingers: (a) transverse view, and (b) side view.

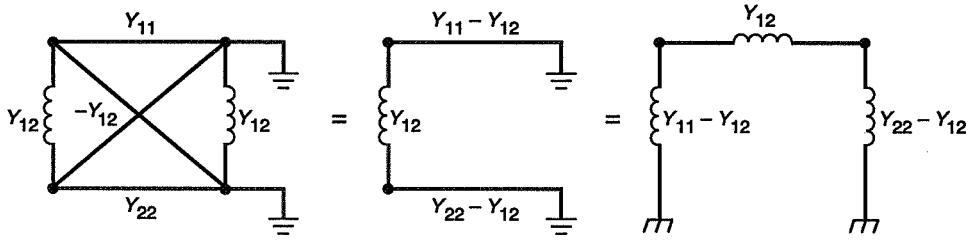


Fig. 17. Equivalent circuit for the homogeneous comb structure.

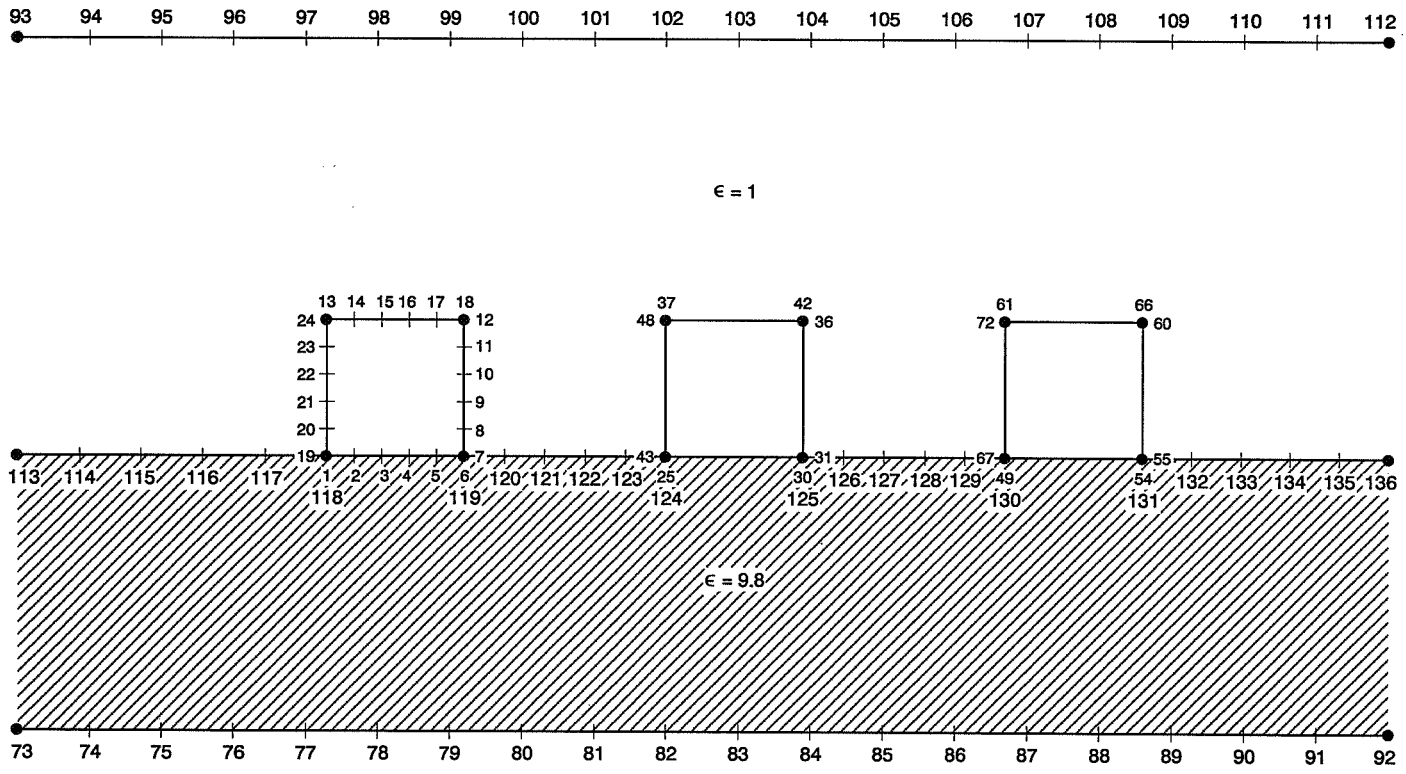


Fig. 18. Location of the nodes used to calculate the capacitance matrix and effective dielectric constant of the partially loaded comb structure.

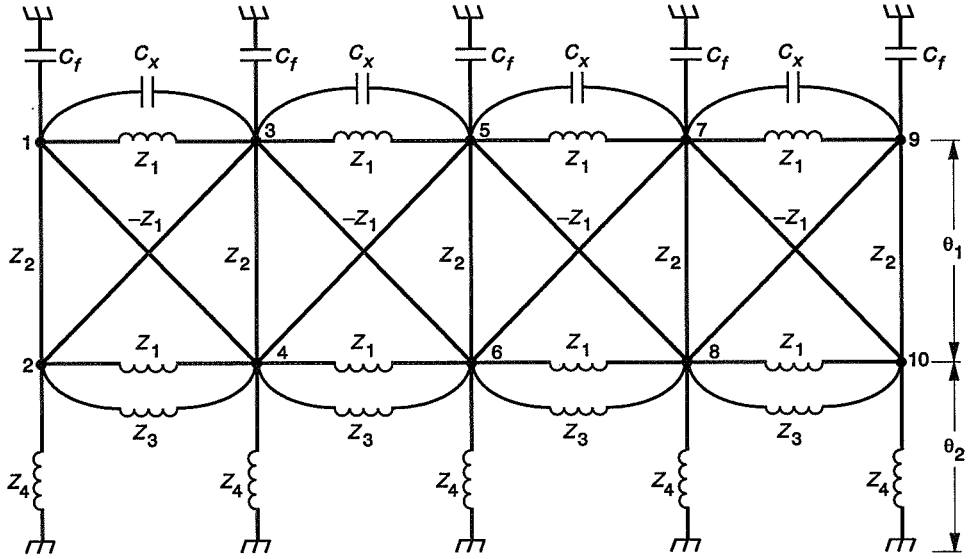


Fig. 19. Equivalent circuit used to model the S-band slow-wave structure of Fig. 16.

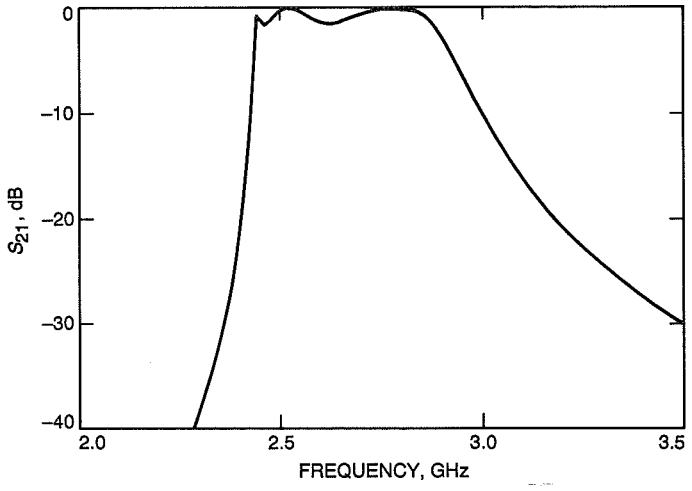


Fig. 20. Calculated S_{21} response for the circuit of Fig. 19.

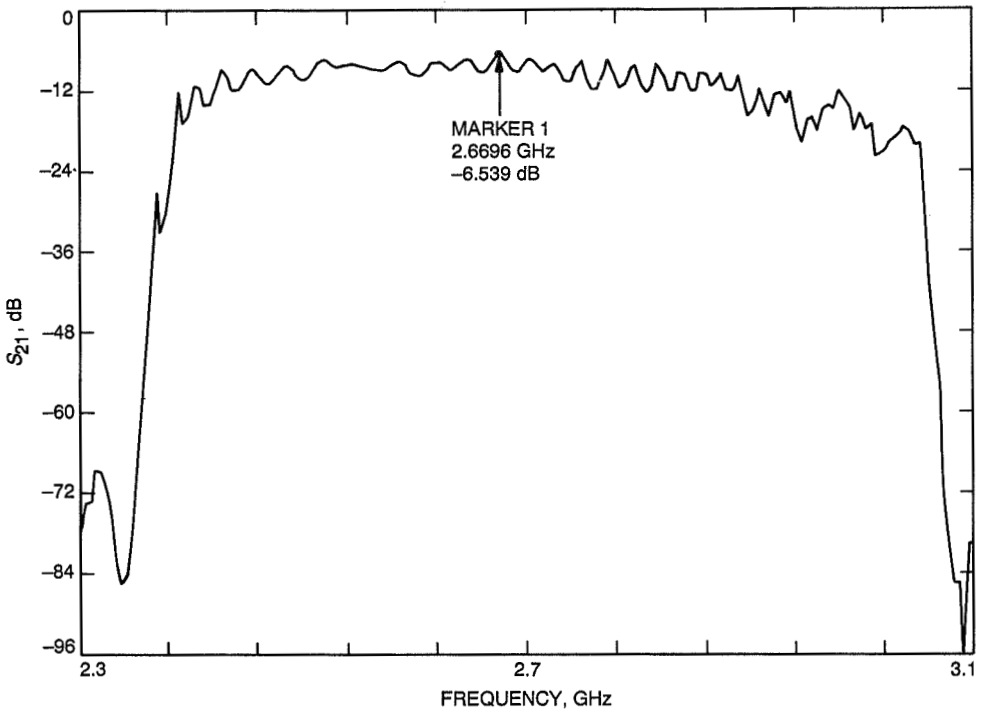


Fig. 21. Measured S_{21} response for the S-band slow-wave structure, which contains 48 fingers.

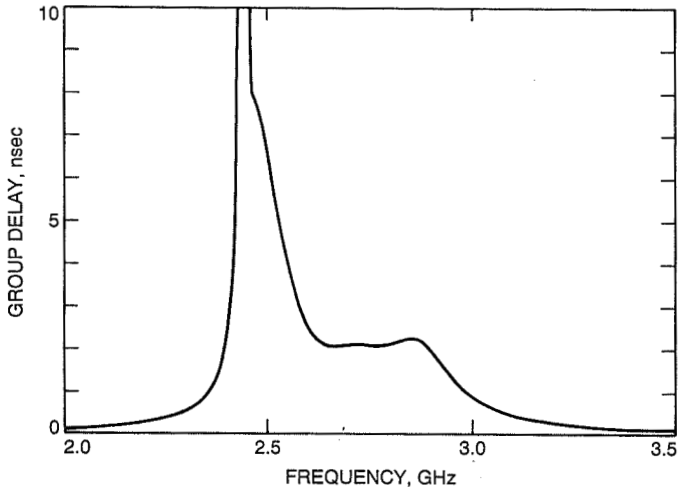


Fig. 22. Calculated group delay for the circuit of Fig. 19.

C-2

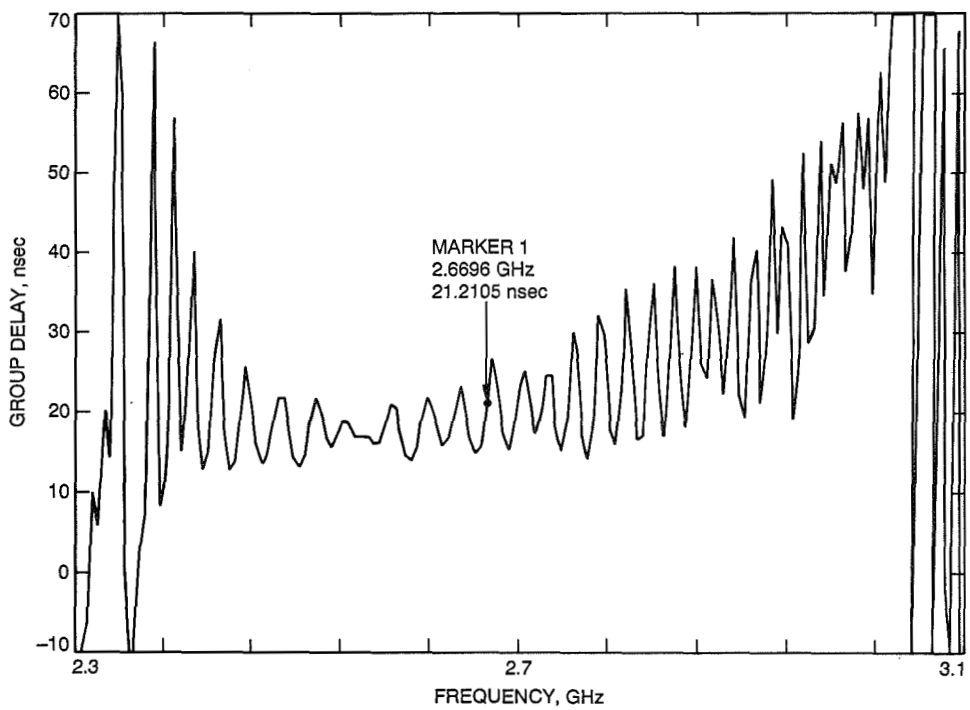


Fig. 23. Measured group delay for the S-band slow-wave structure.

Appendix A

Additional Equivalent Circuits

Several different equivalent circuits for coupled lines using transmission lines have been derived by other researchers. The circuits of Seviora and Sato and Cristal have already been mentioned. The circuit of Seviora can be modified by using the capacitance matrix transformation to yield another circuit, which is discussed by Malherbe [18]. The circuit is shown in Fig. A-1, and in it,

$$Z_1 = \frac{377}{\sqrt{\epsilon_r}} \left\{ \frac{C_{11}}{\epsilon} + \frac{\left(\frac{C_{12}}{\epsilon}\right)\left(\frac{C_{22}}{\epsilon}\right)}{\frac{C_{12}}{\epsilon} + \frac{C_{22}}{\epsilon}} \right\}^{-1} \tag{A-1}$$

$$Z_2 = \frac{377}{\sqrt{\epsilon_r}} \left\{ \frac{\frac{C_{12}}{\epsilon} + \frac{C_{22}}{\epsilon}}{\left(\frac{C_{12}}{\epsilon}\right)^2} \right\} \tag{A-2}$$

$$n = 1 + \frac{\frac{C_{22}}{\epsilon}}{\frac{C_{12}}{\epsilon}} \tag{A-3}$$

The problem of modeling coupled-line networks in an inhomogeneous dielectric medium has also been investigated. The counterpart of the Jones and Bolljahn study (symmetric conductors and homogeneous dielectric material) for the case of inhomogeneous dielectric material (symmetric conductors) was performed by Zysman and Johnson [19].

Cheng and Edwards [20] have shown that the three-transmission-line network of Seviora, if extended to include quasi-TEM coupling, becomes the four-transmission network shown in Fig. A-2. This circuit is being given to improve the modeling of the partially loaded geometry, such as the S-band example considered in this article.

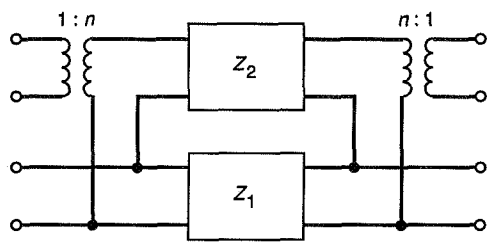


Fig. A-1. Equivalent circuit of Malherbe for coupled lines.

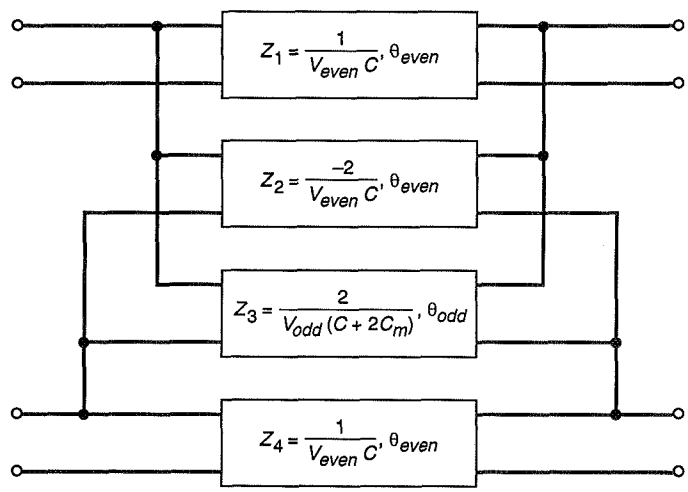


Fig. A-2. Equivalent circuit of Cheng and Edwards to model quasi-TEM coupled lines.

Appendix B

TOUCHSTONE Files

The TOUCHSTONE files used for the X-band and S-band examples described in Section VII are shown in Figs. B-1 and B-2 for reference.

```
DIM

  FREQ GHZ
  RES OH
  IND NH
  CAP PF
  LNG MIL
  TIME NS
  COND /OH
  ANG DEG

VAR
        ZLIN=55.5       !LINE IMPEDANCES CALCULATED FROM CSELF
        ZCRS=-247       !LINE IMPEDANCES CALCULATED
        ZIND=247        !FROM CMUTUAL
        ELIN=150        !ESTIMATED LINE LENGTH
        FLIN=8.0        !AT 8.0 GHZ
        CF1=.1          !FRINGING CAPACITANCE FROM FINGER TIP TO GROUND
        CX1=-.005       !FRINGING CAPACITANCE FROM FINGER TO FINGER

EQN
CKT
        RES   1    0     R=350      !INPUT MATCHING NETWORK
        IND   1    0     L=45
        DEF1P 1          ZIMP

        TLIN  1    2     Z^ZLIN   E^ELIN   F^FLIN
        TLIN  3    4     Z^ZLIN   E^ELIN   F^FLIN
        TLIN  5    6     Z^ZLIN   E^ELIN   F^FLIN
        TLIN  7    8     Z^ZLIN   E^ELIN   F^FLIN
        TLIN  9    10    Z^ZLIN   E^ELIN   F^FLIN
        TLIN  1    4     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  2    3     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  4    5     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  3    6     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  5    8     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  6    7     Z^ZCRS   E^ELIN   F^FLIN
        TLIN  7    10    Z^ZCRS   E^ELIN   F^FLIN
        TLIN  8    9     Z^ZCRS   E^ELIN   F^FLIN
        TLSC  1    3     Z^ZIND   E^ELIN   F^FLIN
        TLSC  2    4     Z^ZIND   E^ELIN   F^FLIN
        TLSC  3    5     Z^ZIND   E^ELIN   F^FLIN
        TLSC  4    6     Z^ZIND   E^ELIN   F^FLIN
        TLSC  5    7     Z^ZIND   E^ELIN   F^FLIN
        TLSC  6    8     Z^ZIND   E^ELIN   F^FLIN
        TLSC  7    9     Z^ZIND   E^ELIN   F^FLIN
        TLSC  8    10    Z^ZIND   E^ELIN   F^FLIN

        CAP   1    0     C^CF1
        CAP   2    0     C^CF1
        CAP   3    0     C^CF1
```

Fig. B-1. TOUCHSTONE file to model the X-band slow-wave structure.

```
        CAP    4     0      C^CF1
        CAP    5     0      C^CF1
        CAP    6     0      C^CF1
        CAP    7     0      C^CF1
        CAP    8     0      C^CF1
        CAP    9     0      C^CF1
        CAP    10    0      C^CF1
        CAP    1     3      C^CX1
        CAP    2     4      C^CX1
        CAP    3     5      C^CX1
        CAP    4     6      C^CX1
        CAP    5     7      C^CX1
        CAP    6     8      C^CX1
        CAP    7     9      C^CX1
        CAP    8     10     C^CX1
        DEF2P  1     9      XTWM

TERM
        XTWM     ZIMP     ZIMP

OUT
        XTWM      DB[S21]  GR1     !PLOTS LOG MAGNITUDE OF S21
        XTWM      DB[S11]  GR2     !PLOTS LOG MAGNITUDE OF S11
        XTWM      TD[S21]  GR3     !PLOTS GROUP DELAY

FREQ
   SWEEP    6.5       9.5       .02

GRID
   RANGE    6.5       9.5      0.5

   GR1    0   -50    10
   GR2    0   -30     5
   GR3    0     5    0.5
```

Fig. B-1 (contd)

```
DIM
  FREQ GHZ
  RES OH
  IND NH
  CAP PF
  LNG MIL
  TIME NS
  COND /OH
  ANG DEG

VAR
        CF1=.0214     !FRINGING CAPACITANCE FROM FINGER TIP TO GROUND
        CX1=.065      !FRINGING CAPACITANCE FROM FINGER TO FINGER
        ELN1=16.5     !LINE LENGTH OF UNLOADED PORTION OF THE FINGER
        ELN2=56       !LINE LENGTH OF LOADED PORTION OF FINGER
        FLIN=2.0      !AT 2.0 GHZ

EQN


CKT
        RES     1   0   R=900
        IND     1   0   L=500
        DEF1P   1       ZIMP

        TLSC    1   3   Z=346      E^ELN1    F^FLIN
        TLIN    1   2   Z=184      E^ELN1    F^FLIN
        TLIN    1   4   Z=-346     E^ELN1    F^FLIN
        CAP     1   0   C^CF1
        CAP     1   3   C^CX1
        TLSC    2   4   Z=346      E^ELN1    F^FLIN
        TLSC    2   4   Z=236      E^ELN2    F^FLIN
        TLSC    2   0   Z=140      E^ELN2    F^FLIN
        TLIN    2   3   Z=-346     E^ELN1    F^FLIN
        TLIN    3   4   Z=184      E^ELN1    F^FLIN
        CAP     3   0   C^CF1
        TLSC    3   5   Z=346      E^ELN1    F^FLIN
        CAP     3   5   C^CX1
        TLIN    3   6   Z=-346     E^ELN1    F^FLIN
        TLSC    4   0   Z=140      E^ELN2    F^FLIN
        TLSC    4   6   Z=346      E^ELN1    F^FLIN
        TLSC    4   6   Z=236      E^ELN2    F^FLIN
        TLIN    4   5   Z=-346     E^ELN1    F^FLIN
        TLIN    5   6   Z=184      E^ELN1    F^FLIN
        CAP     5   0   C^CF1
        TLIN    5   8   Z=-346     E^ELN1    F^FLIN
        TLSC    5   7   Z=346      E^ELN1    F^FLIN
        CAP     5   7   C^CX1
        TLSC    6   0   Z=140      E^ELN2    F^FLIN
        TLSC    6   8   Z=346      E^ELN1    F^FLIN
        TLSC    6   8   Z=236      E^ELN2    F^FLIN
        TLIN    6   7   Z=-346     E^ELN1    F^FLIN
```

Fig. B-2. TOUCHSTONE file to model the S-band slow-wave structure.

```
        TLIN    7    8    Z=184       E^ELN1    F^FLIN
        CAP     7    0    C^CF1
        TLSC    7    9    Z=346       E^ELN1    F^FLIN
        CAP     7    9    C^CX1
        TLIN    7    10   Z=-346      E^ELN1    F^FLIN
        TLSC    8    0    Z=140       E^ELN2    F^FLIN
        TLSC    8    10   Z=346       E^ELN1    F^FLIN
        TLSC    8    10   Z=236       E^ELN2    F^FLIN
        TLIN    8    9    Z=-346      E^ELN1    F^FLIN
        CAP     9    0    C^CF1
        TLIN    9    10   Z=184       E^ELN1    F^FLIN
        TLSC    10   0    Z=140       E^ELN2    F^FLIN
        DEF2P   2    10   STWM2

TERM

     STWM2    ZIMP   ZIMP


OUT
         STWM2   DB[S21]   GR1      !PLOTS LOG MAGNITUDE OF S21
         STWM2   DB[S11]   GR2      !PLOTS LOG MAGNITUDE OF S11
         STWM2   TD[S21]   GR3      !PLOTS GROUP DELAY

FREQ
   SWEEP     2.0      3.5         0.02

GRID
   RANGE     2.0      3.5         0.50

   GR1     0    -40     10
   GR2     0    -30     5
   GR3     0     10     1
```

Fig. B-2 (contd)

References

Some of the articles listed below can be found in the two following reprint collections of IEEE articles dealing with coupled line circuits.

Leo Young, ed., *Microwave Filters Using Parallel Coupled Lines*, Dedham, Massachusetts: Artech House, Inc., 1972 (contains 30 articles).

Leo Young, ed., *Parallel Coupled Lines and Directional Couplers*, Dedham, Massachusetts: Artech House, Inc., 1972 (contains 26 articles).

[1] R. Seviora, "Equivalent Circuit for Parallel Conductor Array," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-16, no. 10, pp. 875–877, October 1968.

[2] R. Sato and E. G. Cristal, "Simplified Analysis of Coupled Transmission Line Networks," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-18, no. 3, pp. 122–131, March 1970.

[3] H. J. Riblet, "An Explicit Derivation of the Relationships Between the Parameters of an Interdigital Structure and the Equivalent Transmission Line Cascade," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-15, no. 3, pp. 161–166, March 1967.

[4] E. M. T. Jones and J. T. Bolljahn, "Coupled-Strip Transmission Line Filters and Directional Couplers," *I.R.E. Trans. on Microwave Theory and Techniques*, vol. MTT-4, no. 2, pp. 75–81, April 1956.

[5] R. J. Wenzel, "The Modern Network Theory Approach to Microwave Filter Design," *IEEE Trans. on Electromagnetic Compatibility*, vol. EMC-10, no. 2, pp. 196–209, June 1968.

[6] R. J. Wenzel, "Exact Design of TEM Microwave Networks Using Quarter Wave Lines," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-12, no. 1, pp. 94–111, January 1964.

[7] R. J. Wenzel, "Exact Theory of Interdigital Band-Pass Filters and Related Coupled Structures," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-13, no. 5, pp. 559–575, September 1965.

[8] R. J. Wenzel, "Theoretical and Practical Applications of Capacitance Matrix Transformations to TEM Network Design," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-14, no. 12, pp. 635–647, December 1966.

[9] A. Djordjevic, R. F. Harrington, T. Sarkar, and M. Bazdar, *Matrix Parameters for Multiconductor Transmission Lines: Software and Users Manual,* Dedham, Massachusetts: Artech House, Inc., 1989.

[10] W. J. Getsinger, "Coupled Rectangular Bars Between Parallel Plates," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-10, no. 1, pp. 65–72, January 1962.

[11] TOUCHSTONE, EEsof, Inc., 1987.

[12] S. B. Cohn, "Shielded Coupled-Strip Transmission Line," *IRE Trans. on Microwave Theory and Techniques*, vol. MTT-3, no. 5, pp. 29–38, October 1955.

[13] H. E. Green, "The Numerical Solution of Some Important Transmission-Line Problems," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-13, no. 5, pp. 676–692, September 1965.

[14] R. E. Diaz, "The Discrete Variational Conformal Technique for the Calculation of Strip Transmission-Line Parameters," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-34, no. 6, pp. 714–722, June 1986.

[15] D. W. Kammler, "Calculation of Characteristic Admittances and Coupling Coefficients for Strip Transmission Lines," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-16, no. 1, pp. 925–937, November 1968.

[16] C. Wei, R. F. Harrington, J. R. Mautz, and T. K. Sarkar, "Multiconductor Transmission Lines in Multi-layered Dielectric Media," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-32, no. 4, pp. 439–450, April 1984.

[17] S. M. Perlow, "Analysis of Edge-Coupled Shielded Strip and Slabline Structures," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT-35, no. 5, pp. 522–529, May 1987.

[18] J. A. G. Malherbe, *Microwave Transmission Line Filters,* Dedham, Massachusetts: Artech House, Inc., 1979.

[19] G. Zysman and A. K. Johnson, "Coupled Transmission Line Networks in an Inhomogeneous Dielectric Medium," *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-17, no. 10, pp. 753–759, October 1969.

[20] S. Cheng and M. L. Edwards, "TEM Equivalent Circuits for Quasi-Tem Couplers," *1990 IEEE MTT-S International Microwave Symposium Digest*, vol. 1, IEEE, Piscataway, New Jersey, p. 387, May 1990.

Spectral Filters for Laser Communications

K. Shaik
Communications Systems Research Section

Optical communication systems must perform reliably under strong background light interference. Since the transmitting lasers operate within a narrow spectral band, high signal-to-noise ratios can be achieved when narrowband spectral optical filters can be used to reject out-of-band light. This article develops a set of general requirements for such filters and provides an overview of suitable spectral filter technologies for optical communication systems.

I. Introduction

Optical communication systems will be required to function under diverse, sometimes hostile, channel-noise conditions. Various techniques and technologies will be employed to reject background light (optical noise), while, at the same time, providing maximum possible signal throughput efficiency. Spectral optical filters (SOFs) in this regard will play an important role in determining the system sensitivity.

An SOF is usually designed for a given center wavelength, λ_c, at which the filter provides maximum throughput efficiency, η. The throughput efficiency decreases as the incident wavelength moves away from λ_c. Wavelengths $\lambda_c \pm \Delta\lambda/2$, where the filter efficiency drops to 0.5 of its peak efficiency, defines the filter bandwidth, $\Delta\lambda$. These characteristics are critical to the performance of the SOF. The bandwidth of the filter determines how well it blocks out-of-band radiation from reaching the detector. The spectral bandwidth of the SOF should be as small as possible, especially if the detection process is limited by the background optical noise. Fortunately, spectral bandwidths of lasers, which are to be used as transmitters in optical communication systems, are extremely narrow. Narrowband SOFs tuned to the central frequency of the transmitter laser can provide the necessary signal-to-noise ratio (SNR) under high-background conditions. However, due to possibly large, relative velocities between the receiver and the transmitter, Doppler shifts as large as 0.1 nm are possible. Hence, the narrowband SOFs must have the ability to tune their center frequency over this range.

Narrowband SOFs usually come with high insertion loss, which makes them unsuitable when noise background is low. For example, deep-space optical communication systems operating during the nighttime will be signal limited. Under such channel conditions, it will be important to obtain the highest possible signal throughput efficiencies, even at the expense of increasing the passband. As the bandwidth of the SOF is increased, the requirement on its tunability to counter Doppler shift becomes less critical.

Field of view (FOV) of an SOF is also an important parameter. For most filter systems the center wavelength shifts away from the design value with the angle of arrival of the signal beam at the filter. This results in a decrease in the filter throughput efficiency as the field angle increases. Other important properties of SOFs include clear aperture size and stability. Restrictions on both parameters arise from limitations on availability of suitable materials and manufacturability.

As stated earlier, optical noise can be very severe when the communication system has to operate within a few degrees of the sun (daytime), or relatively benign when operating during the nighttime. It is not expected that a single SOF system will be able to cover the whole range of possible channel noise conditions to provide optimal SNR for optical communication applications. Instead, several SOF systems, each optimized for a set of typical channel conditions will most likely have to be developed. The need for a narrowband SOF for use under high-background, optical-noise conditions, and a relatively broad bandwidth SOF with low insertion loss for use under dark background conditions has been identified.

To summarize, the fundamental requirements for these SOFs for optical communications will include the following:

(1) The narrowband SOFs for operation under intense background optical noise shall have
   (a) a passband of less than 0.05 nm
   (b) an overall transmission efficiency higher than 0.2
   (c) central wavelength tunability over 0.2 nm
   (d) central wavelength stability within 10 percent of the filter bandwidth

(2) The wideband SOFs for operation under low-background, optical-noise conditions shall have
   (a) a passband between 0.1 to 1 nm
   (b) a transmission efficiency of about 0.7
   (c) central wavelength tunability, as appropriate, to eliminate transmission loss due to Doppler shift
   (d) central wavelength stability within 10 percent of the filter bandwidth

(3) The field of view shall be as large as the design value for the receiver optics to minimize off-axis insertion loss and change in the center frequency.

(4) The clear aperture shall be as large as the design beam diameter for the optical communication instrument.

(5) The center wavelength of the SOFs shall match the center wavelength of the transmitter laser.

(6) Filter surfaces shall have negligible contribution to the system blur diameter.

Section II provides an overview of atomic resonance, birefringent, Fabry-Perot, Faraday, and other SOF technologies as well as information on manufacturability of such filters. In Section III, available SOF technologies and their relative merits for use with optical communication systems are discussed.

II. SOF Technologies

Several SOF technologies have been developed which may be suitable for optical communication systems. A discussion of such technologies and their capabilities is given below.

A. Atomic Resonance Filters

An atomic resonance filter (ARF) consists of an atomic vapor cell sandwiched between two conventional high-efficiency optical bandpass filters [1–6]. The input bandpass filter is centered at the input signal wavelength from the laser transmitter $\lambda_c(i)$. Light at this wavelength passes into the vapor cell and is absorbed. The atoms re-emit light at a shifted wavelength $\lambda_c(o)$. The output bandpass filter centered at $\lambda_c(o)$ allows the light to pass through, where it is then detected by a photodetector. If $\lambda_c(i)$ and $\lambda_c(o)$ differ from each other by several nanometers, relatively common SOFs can be used as input and output filters that have high throughput efficiencies and nonoverlapping passbands. Hence, the effective ARF bandwidth, which is essentially determined by the Doppler line width of the atomic absorption, is extremely narrow.

ARFs are based on atomic transitions and as such are insensitive to temperature fluctuations. Also, the vapor cells are insensitive to the angle of arrival of the signal light, i.e., ARFs possess a 2π-steradian FOV. The presence of input and output bandpass filters, however, will restrict the acceptance angle of the filter system to a few degrees.

ARFs operate at numerous discrete wavelengths. Table 1 lists some of the developed ARFs, with their operating wavelengths in the visible region [1].

The response time of an ARF is an important consideration, as it will determine the upper limit on the information rate. Typical rapid spontaneous emission times are ~30 nsec, which will allow an information throughput of about 30 MHz. However, it is possible to introduce a quenching gas, usually a noble gas, to decrease the response time, but at the expense of an increased passband. With this procedure, information rates in the gigahertz range are possible.

Filter conversion efficiency for the ARFs is defined as

$$\eta_c = \frac{N_o}{N_i} \tag{1}$$

where N_o is the number of re-emitted photons at $\lambda_c(o)$ and N_i is the number of incident photons at $\lambda_c(i)$. It can be as high as 95 percent, as in the case of the Ca resonance filter at 432 nm [2]. For the Rb resonance filter operating at frequency-doubled Nd:YAG wavelength (532 nm), $\eta_c = 0.16$. However, with a vapor-cell design that traps 532-nm photons inside, η_c for the Rb filter can be increased to 0.28 [5]. For a complete Rb ARF filter system, which will include the input and output bandpass filters, the overall throughput efficiency, η, will drop to about 0.15.

The minimum ARF bandwidth is determined by the Doppler line width of the atomic absorption. Typically, this bandwidth is about 0.001 nm, which corresponds to about 1 GHz in the visible. However, optical communication systems may experience Doppler shifts as high as 50 GHz, due to relative motion between the receiver and the transmitter. This problem can be solved by introducing an inert buffer gas into the chamber. Collisions between the atomic vapors and the buffer gas can be used to broaden the passband of the filter beyond the Doppler shift. Also, the central absorbing wavelength changes with the buffer gas pressure, providing a mechanism for tuning the ARFs [3].

B. Birefringent Filters

Lyot-Ohman and Solc filters, two of the most commonly used birefringent filters, use an array of birefringent crystals and polarizers to select a narrow passband [7–11]. Typical materials employed for such filters are quartz and calcite.

A Lyot-Ohman filter consists of several crystal plates whose thicknesses vary geometrically, each twice as thick as its predecessor. The plates are separated by polarizing elements. Each birefringent crystal has a transmission spectrum given by

$$\tau = \cos^2[\pi \Delta n \; d/\lambda] \tag{2}$$

where Δn is the birefringence of the element, d is the thickness of the crystal, and λ is the wavelength. The overall transmission spectrum of the filter is obtained by multiplying the transmission spectrum of each of the crystal elements. The bandwidth of the Lyot-Ohman filter at full-width-half-maximum (FWHM) is

$$\Delta\lambda = 0.44\lambda_c^2/(2^{N-1}\Delta n \; d_0) \tag{3}$$

and the free spectral range (FSR), i.e., the spectral distance between adjacent transmission maxima is approximately

$$FSR \simeq \lambda_c^2/(\Delta n \; d_0) \tag{4}$$

where N is the number of crystal elements, d_0 is the thickness of the thinnest element, and λ_c is the central wavelength.

A Solc, or lossless, birefringent filter varies the angular orientation of the crystals rather than their thickness to achieve narrowband filtering. It consists of a series of birefringent plates of equal thickness between a single pair of polarizers. For an N-element filter, the orientation of the kth plate is given by

$$\phi_k = (2k-1)\rho \tag{5}$$

with reference to the input polarizer axis where $\rho = 45/N$ deg. The transmission spectrum for this design is given by

$$\tau = \left[\frac{\sin N\beta}{\sin\beta}\cos\beta \; \tan\rho\right]^2 \tag{6}$$

where $\cos\beta = \cos\rho \; \cos(\pi\Delta n d/\lambda)$.

More recent designs based on Solc filters are rugged, simpler to build, and have greater spectral versatility. These filters have the ability to place the passband at an arbitrary wavelength and can be tuned over a broad range. The transmission profile and the bandwidth can also be engineered with greater control. Rotating waveplate tuning has been used on a space-qualified, 0.005-nm bandwidth filter, which can be tuned to an accuracy of about 0.0005 nm [7].

C. Fabry-Perot Filters

The Fabry-Perot interferometer (FPI) consists of two parallel, flat, and transparent plates coated for high reflectivity and made of low-absorption materials[1] [11–17].

[1] Product literature, Barr Associates.

The space between the two plates forms a cavity that is resonant at specific wavelengths determined by the optical thickness of the gap. The FPI transmits a narrow spectral band at each of a series of wavelengths λ, which are given by

$$m\lambda = 2nd\cos\phi \tag{7}$$

where m is the order of interference, n is the refractive index of the medium in the gap, d is the thickness of the gap, and ϕ is the angle of incidence within the cavity. The transmission of an FPI is given by

$$\tau(\lambda,\phi) = \frac{T}{(1-R)^2}\left[1+\frac{4R}{(1-R)^2}\sin^2\left(\frac{2\pi nd\cos\phi}{\lambda}\right)\right]^{-1} \tag{8}$$

where T and R are the intensity transmission and reflection coefficients, respectively, of the coatings, which are assumed to be identical.

The total usable, angular field 2ϕ of an FPI is determined by the allowable shift $\delta\lambda$ in the central transmission wavelength across the field. It is given by

$$\phi = [2\delta\lambda/\lambda]^{1/2} \tag{9}$$

measured in radians. If the allowable shift in center wavelength is $\delta\lambda = 0.025$ nm, and if $\lambda = 532$ nm, the resulting FOV of the FPI is found to be about 20 mrad.

The FPI can be tuned over a wide range by tilting it with respect to the incoming optical beam. For a tilt of magnitude $\delta\phi$, the shift in the central wavelength is $\delta\lambda = -(2nd/m)\sin\phi\ \delta\phi$.

Dielectric, thin film interference filters are the most abundant form of Fabry-Perot filters (FPFs). Multilayers of dielectric materials, such as cryolite and zinc sulfide, with differing refractive indices are deposited alternately on a substrate, usually glass. Thicknesses and the number of the alternating thin layers are engineered to allow peak transmission at the desired central wavelength λ_c and bandwidth $\Delta\lambda$. The passband range for interference filters is 0.2 to 50 nm. Interference filters with narrower bandwidths are possible by depositing increasingly larger numbers of dielectric layers. This makes such filters difficult to manufacture and quite fragile. Studies of narrowband interference filters have shown that the peak transmission wavelength drifts unpredictably to shorter wavelengths in time due to thermal and radiation shock [16]. Typical transmission efficiency for a 1-nm bandwidth interference filter is about 0.7.

FPFs with a thick solid cavity, instead of interference filters, can be manufactured for passbands less than 0.1 nm. The reflecting layers are deposited on either side of a transparent substrate, usually made from a highly polished fused silica disk. Optical thickness of the substrate must be uniform to better-than-$\lambda/100$. Silica disks of thicknesses between 50 and 1000 μm, corresponding to a range of passbands between 0.1 and 0.005 nm, respectively, can be manufactured. The thicker disks are easier to manufacture. FPFs with bandwidth $\Delta\lambda = 0.03$ nm, diameter $D = 75$ mm, and transmission efficiency $\tau = 0.7$ have been reported [14]. However, an input passband filter is required to remove adjacent transmission channels, which reduces the overall system efficiency to about 0.5.

To obtain still narrower bandwidths, two solid etalons must be used in series. The second cavity is necessary to suppress the transmission channels adjacent to the desired λ_c, which are allowed to pass through the primary etalon. With this technique, it is possible to construct 0.005-nm filters with $\tau \simeq 0.2$, which employ two solid etalons and a conventional interference filter. By using a combination of tilt and temperature control, such a filter can be tuned over a range of 3 nm.

D. Faraday and Other Types of Filters

The Faraday anomalous dispersion optical filter (FADOF) has been discovered recently [18–19]. FADOFs are based entirely on the resonant Faraday effect. With the Faraday effect, the polarization of the incident light rotates as it passes through an active medium in the presence of a magnetic field. The relation between the angle α of polarization rotation and the applied magnetic field H is

$$\alpha = VHl\cos\psi \tag{10}$$

where V is the Verdet constant for the active medium, l is the interaction length through the medium, and ψ is the angle between the direction of the magnetic field and the direction of propagation of the light beam.

The filter system consists of a vapor cell sandwiched between two crossed polarizers. When the applied magnetic field, the vapor density, and the length of the cell are properly adjusted, the polarization of only a narrow band of frequencies is rotated by 90 deg. This narrow band of frequencies is then transmitted by the output polarizer.

Unlike the ARF, the FADOF is an imaging filter, i.e., it does not change the spatial distribution of incident energy.

The bandwidth for the FADOF is limited by the Doppler line width of the atomic vapor, which is about 0.001 nm in the visible. The response time is also limited essentially by the Fourier transform of the Doppler bandwidth, which is about 1 nsec for the visible region. The FADOF, like the ARF, has a wide FOV.

In a laboratory setup the peak throughput efficiency of the FADOF system, which consists of an Rb vapor cell and two crossed polarizers tuned to a center frequency $\lambda_c = 780$ nm, has been shown to be about 0.63. Theoretical models predict throughput efficiencies higher than 0.9. The ability to track Doppler shifts in the signal due to relative motion between the transmitter and the receiver is being investigated.

A number of unconventional concepts in optical spectral filters using photorefractive, acousto-optic, and electro-optic effects are currently under investigation. However, the technology for such filters has been unable to produce narrowband SOFs with reasonable efficiency or is simply not yet mature enough for use with the optical communication systems.

III. Discussion

For comparison and ready reference, various filter capabilities have been summarized in Table 2. It is assumed that the center frequency for the compared filters is 532 nm, except for the FADOF, for which the results for a 780-nm filter are listed. The bandwidth, efficiency, and clear aperture columns in the table show typical values. The filter efficiencies shown here are typical of a complete SOF system. For example, the efficiency of a single cavity FPF can be as high as 0.7, but for a complete system, which must include a bandpass filter to reject adjacent transmission bands, the transmission efficiency drops to about 0.5. The FOV is calculated for a shift in the center frequency equal to half the listed bandwidth for each of the filters at 532 nm.

Typical dielectric interference filters can provide 0.2-nm bandwidths with throughput efficiency $\eta \simeq 0.7$. To obtain bandwidths $\Delta\lambda \leq 0.1$ nm, solid cavity FPFs are used. Such filters must be placed in a temperature-controlled environment. This temperature-controlled environment can, however, be used to tune the central wavelength of the passband at a typical rate of 0.02 nm/deg C in the visible range. With active temperature control, FPFs with a single cavity can provide bandpass as small as 0.05 nm in the visible, with $\eta \simeq 0.5$. The peak throughput efficiency for double cavity FPFs, however, falls to about 0.2 for bandwidths as small as 0.005 nm. Another important limitation on the double cavity FPFs relates to the FOV, which is on the order of a few mrad. Currently 0.05-, 0.2-, and 1-nm bandpass filters can be routinely fabricated up to dimensions of 3, 5, and 15 cm, respectively [17]. Larger sizes are available by custom processing.

Birefringent filters provide an attractive alternative when narrow bandpass filters are required. Such filters, with bandpasses between 0.04 to 0.3 nm, are possible with 0.25-transmission efficiency. Aperture sizes of 7 cm can be fabricated.

The ARFs provide extremely narrow passbands with reasonable efficiencies, and, unlike the FPFs and the birefringent filters, they do not have a strong restriction on the FOV. The ARFs provide greatly improved performance, but only for background-limited detection. Hence, this is the only case that warrants the additional complexity associated with ARF use, as compared with conventional optical filters.

The newly discovered FADOF has wide FOV and narrow bandwidth comparable to the ARF. The FADOF throughput efficiency is expected to be much higher. Also, the FADOF can be used in applications where the preservation of the incident image field is necessary.

Figure 1 graphically shows the relationship between the bandwidth, $\Delta\lambda$, and the overall throughput efficiency, η, for the various filters discussed above. Figure 2 shows the relationship between $\Delta\lambda$ and the filter system FOV. In either of the two figures, a good candidate SOF for optical communications will lie in the upper left quadrant. It seems that the FADOF can potentially be an important addition to the SOF technologies for optical communications. However, more work needs to be done to resolve questions on its tunability and to develop a viable and robust filter. The double cavity FPFs may not be able to provide the necessary FOV, and may have to be dropped in favor of the ARFs when the signal is limited by the background noise. Cheaper single cavity FPFs, birefringent filters, or interference filters may become useful when wider bandwidths can be tolerated or are desired.

References

[1] J. A. Gelbwachs, "Atomic resonance filters," *IEEE Jour. Quant. Elect.*, vol. 24, pp. 1266–1277, 1988.

[2] J. A. Gelbwachs, "Active wavelength shifting in atomic resonance filters," *IEEE Jour. Quant. Elect.*, vol. 26, pp. 1140–1147, 1990.

[3] J. A. Gelbwachs and M. D. Tabat, "Solar background rejection by a pressure broadened atomic resonance filter operating at a Fraunhofer wavelength," *Opt. Let.*, vol. 14, pp. 211–213, 1989.

[4] T. M. Shay and D. F. Garcia, "Theoretical model for a background noise limited laser-excited optical filter for doubled Nd lasers," *IEEE Jour. Quant. Elect.*, vol. 26, pp. 1135–1139, 1990.

[5] T. M. Shay and Y. C. Chung, "Ultrahigh-resolution, field-of-view optical filter for the detection of doubled Nd:YAG radiation," *Opt. Let.*, vol. 13, pp. 443–445, 1988.

[6] J. A. Gelbwachs, "422.7-nm atomic filter with superior solar background rejection," *Opt. Let.*, vol. 15, pp. 236–238, 1990.

[7] W. J. Rosenberg and A. M. Title, "Solc filter engineering," *SPIE*, vol. 307, pp. 106–111, 1981.

[8] W. G. Driscoll, *Handbook of Optics*, New York: McGraw-Hill, pp. 8-111–8-115, 1978.

[9] I. Solc, "Birefringent chain filters," *J. Opt. Soc. Am.*, vol. 55, p. 621, 1965.

[10] A. M. Title and W. J. Rosenberg, "Tunable birefringent filters," *Opt. Eng.*, vol. 20. p. 815, 1981.

[11] W. J. Gunning, "Electro-optically tuned spectral filters: a review," *Opt. Eng.*, vol. 20, pp. 837–845, 1981.

[12] J. V. Ramsey, "Very narrow band interference filters," *Proceedings of the ASA*, vol. 2, pp. 66–68, November 1967.

[13] P. D. Atherton, N. K. Reay, J. Ring, and T. R. Hicks, "Tunable Fabry-Perot filters," *Opt. Eng.*, vol. 20, pp. 806–814, 1981.

[14] A. M. Title, "Fabry-Perot interferometers as narrowband optical filters," in *New techniques in space astronomy, Proceedings of the symposium*, Munich, West Germany, pp. 325–332, August 10–14, 1970.

[15] D. R. Gibson and P. H. Lissberger, "Optical properties of narrowband spectral filter coatings related to layer structure and preparation," *Appl. Opt.*, vol. 22, pp. 269–281, 1983.

[16] A. M. Title, T. P. Pope, and J. P. Andelin, Jr., "Drift in interference filters, Part 1," *Appl. Opt.*, vol. 13, pp. 2675–2679, 1974.

[17] P. Baumeister, "Theory of rejection filters with ultranarrow bandwidths," *J. Opt. Soc. Am.*, vol. 71, pp. 604–606, 1981.

[18] T. M. Shay and D. F. Garcia, "*Theoretical model for a Faraday anomalous dispersion optical filter*," paper presented at the International Conf. on Lasers, New Orleans, Louisiana, December 4–8, 1989.

[19] B. Ying and T. M. Shay, "*Theoretical model for a Faraday anomalous dispersion optical filter operating at 423 nm*," paper presented at the Annual meeting of the IEEE Lasers and Electro-Optics Society, Boston, Massachusetts, November 1990.

Table 1. Operating wavelengths for developed ARFs

Atom	$\lambda_c(i)$, nm	$\lambda_c(o)$, nm	Pumping Source
Ca	465, 459	452, 894	Passive
Rb	Over 20 between 487–776 (including doubled Nd:YAG)	420	Diode laser
Mg	516, 517, 518	383	Optical
Tl	535 (doubled Nd:BEL)	378	Photochemical, thermal

Table 2. Filter characteristics comparison chart

Filter	Bandwidth, nm	Efficiency	FOV, rad	Clear Aperture, cm	Tunability	Stability	Complexity	Coverage
Atomic resonance filter	0.001	0.15	0.3	Large	Low	High	High	UV to visible
Birefringent filter	0.04	0.25	0.02	7	High	High	Moderate	UV to near IR
Fabry-Perot filters:								
Interference	1	0.7	0.1	15	High	Low	Low	UV to IR
Single cavity	0.05	0.5	0.02	5	High	High	Moderate	UV to IR
Double cavity	0.005	0.25	0.01	3	High	High	Moderate	UV to IR
Faraday anomalous dispersion optical filter	0.001	0.63	0.3	Large			Moderate	UV to IR

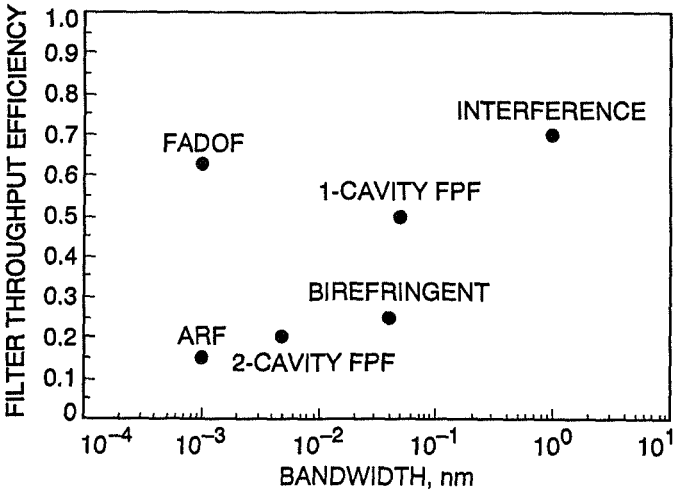


Fig. 1. SOF throughput efficiency versus bandwidth.

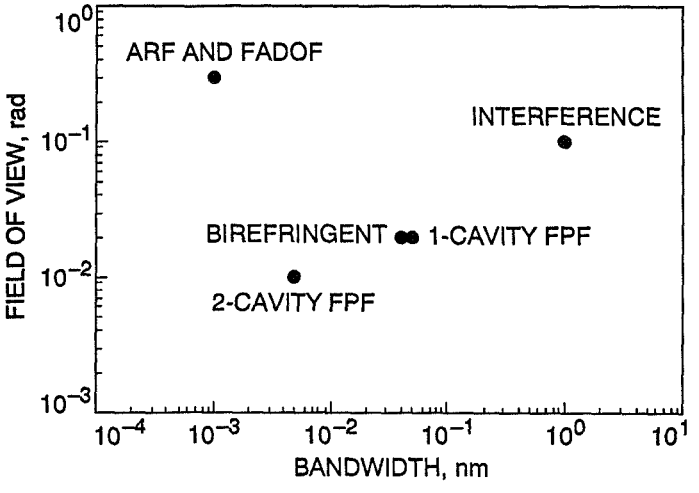


Fig. 2. SOF FOV versus bandwidth.

Design and Testing of an Active Quenching Circuit for an Avalanche Photodiode Photon Detector

D. Arbel
Radio Frequency and Microwave Subsystems Section

J. A. Schwartz
Communications Systems Research Section

The photon-detection capabilities of avalanche photodiodes (APDs) operating above their theoretical breakdown voltages are described, with particular attention given to the needs for and methods of quenching an avalanche once breakdown has occurred. A brief background on the motives of and previous work with this mode of operation is presented. Finally, a description of the design and testing of an active quenching circuit is given. Although the active quenching circuit did not perform as expected, knowledge was gained as to the signal amplitudes necessary for quenching and the need for a better model for the above-breakdown circuit characteristics of the Geiger-mode APD.

I. Introduction

For deep-space optical communications, received data signals will consist of a few photons per bit of received information. One of the enabling technologies for the implementation of optical communications in the deep-space environment is the development of detectors with sufficient sensitivity and gain to operate in a single-photon detection mode. To a degree, conventional photomultiplier tubes (PMTs) are suitable for this purpose, when they are used in ground-based receivers. The signal gain for photomultiplier tubes can be 10^8 for the spectral region of interest (0.5–1.1 μm). But in this range, the quantum efficiency of conventional photomultiplier tubes is limited to ~2 percent at 0.5 μm and less than 1 percent at 1 μm. Out of a need for higher quantum efficiency, high-gain optical detectors, the JPL Optical Communications Group in 1984 began to seek out alternative detectors. This work was undertaken to develop optical-wavelength detectors that would be suitable for use in DSN-type communications applications.

Avalanche photodiodes (APDs) are photosensitive diodes that are operated under a large reverse bias. Incident photons that cause ionization (the creation of hole-electron pairs) within the diode's depletion region, as well as thermally generated carriers, can, with a certain probability, under the influence of a high electric field (~3×10^5 V/cm), start a chain reaction of ionizing colli-

sions that result in the diode breaking down, or avalanching, and conducting current. Additionally, charge carriers trapped within the depletion region during a previous avalanche can also "detrap" and cause another avalanche. In any event, if the reverse bias is less than the breakdown voltage, such a chain reaction is not self-sustaining, and the APD quickly reassumes its nonconducting state until the next ionizing event causes another avalanche. As the reverse bias is increased, both the probability of an ionizing event causing an avalanche and the amount of current resulting from the avalanche, i.e., the gain, increase.

In the early 1980s, researchers at RCA–Canada began experimenting with operating APDs beyond their breakdown voltage to realize photon detectors with signal gains comparable to those of photomultiplier tubes, but with the higher quantum efficiencies associated with APDs [1,2].

If an APD is biased beyond its reverse breakdown voltage, it will suffer avalanche breakdown and begin to conduct current. Unless the current is limited by external circuitry, the breakdown current will increase and destroy the diode. But it is important to note that breakdown results from free minority carriers transiting the diode's depletion region. If such carriers are made unavailable, breakdown cannot be spontaneously initiated, even if the bias is above the nominal breakdown voltage. In fact, the availability of free carriers can be greatly reduced by cooling the diode. For example, RCA-type C30902 silicon APDs can be operated at ~6 V over breakdown, producing $\leq$10 dark counts (avalanches) per sec when cooled to 200 K [3]. An avalanche can now be triggered as a result of a single incident photon generating a hole-electron pair in the depletion region. Such an avalanche, which has a probability of ~0.4 of occurring under the above conditions, can achieve a current gain of 10^8 to 10^9. As a result, when used in this mode, a single photon can be detected with conventional amplifying techniques. This yields a solid-state detector whose gain is comparable to that of a PMT, but whose photon detection probability is 3–4 times higher. This mode of overbiased operation is referred to as the Geiger mode. The rate at which incident photons can be counted depends on how quickly the external circuitry, preparing the APD for the next photon event, can detect and quench the avalanche.

Deep-space optical communications will likely implement pulse position modulation (PPM) data transmission formats. An optical PPM format requires distinguishing a particular time slot that contains a few signal photons plus some noise photons from a time slot with only noise photons. Increasing the signal-to-noise ratio is achieved by shortening the slot times, thus reducing the number of noise photons per slot. But using currently available lasers and modulators, PPM slot times of 10 nsec are realizable. Implementing this technology requires detectors capable of timing resolutions of a few nanoseconds. Beginning in 1984, the Optical Communications Group at JPL worked on using APDs in the Geiger-mode as high-gain, high-quantum-efficiency optical communication detectors [3–5]. Initially, the APD avalanche was detected and quenched by using the passive network shown in Fig. 1. This circuit was capable of quenching the APD in ~6 μsec [4]. Clearly, quenching times of at least three orders of magnitude shorter than those obtainable with the passive quenching circuit are necessary in order to implement PPM with time slots on the order of 10 nsec. C.-C. Chen has shown that receiver error rates degrade rapidly with increasing dead times [6]. Figure 2 shows the transmitted power penalty incurred by a communications system as a function of the detector dead-time when used in a binary PPM system. It can be seen in Fig. 2 that the power penalty incurred to hold to a particular probability of a bit error (PBE) begins to grow rapidly when the detector dead time exceeds one tenth of the PPM slot width (T_s). Quenching circuits that use high-speed active elements have been developed by several groups [7–9], with 20–30-MHz operation being reported. An evaluation of available electronics indicated that 10-nsec quenching times were feasible. Given this, requirements were gathered for an active quenching circuit with a 10-nsec recovery time capable of operating an APD in the Geiger mode up to 5 V over breakdown.

II. Active Quenching Circuit for Geiger-Mode APD

An active circuit was designed as a joint effort of the Optical Communications Group in the Communications Systems Research Section and the New Products Development Group in the Radio Frequency and Microwave Subsystems Section. A schematic of the design is shown in Fig. 3. The design employs high-speed GaAs components. The purpose of the circuit is to detect and quench an avalanche as quickly as possible, thus minimizing the detector's dead time and maximizing the photon count rate. The assumptions used in the design were that the overbias range would be 3–5 V and that the quenching pulse should pull the APD bias just below the breakdown voltage. A replaceable delay line was used to add flexibility in establishing the optimal length of the quenching pulse.

A small bias voltage VN is applied to one end of the resistor $R1$. This voltage determines the steady-state output voltage of the transimpedance amplifier $U1$. The other

end of $R1$ is the crucial node in the circuit that connects to the cathode of the APD and to the output of the quenching pulse switch. When the −4.5-V quenching pulse is applied, the Schottky barrier diode $Z1$ isolates this node from the input of the transimpedance amplifier. The APD, $D1$, is biased beyond breakdown (−175 V at 200 K) by applying a negative voltage to the anode. At steady state, the diode is an open circuit. At the onset of an avalanche, the diode begins to conduct current, which is subtracted from the (internal) feedback resistor of $U1$ and which causes the output voltage to decrease in amplitude (but increase in absolute value, since it is negative). The output of the transimpedance amplifier is connected to the negative input of the comparator, $U2$, through $R2$. The positive input of $U2$ is connected through $R3$ to a variable voltage reference, which is set 30–100 mV above the nominal value of the negative input. $R2$ and $R3$ help reduce oscillations of the comparator that result from the fact that the two inputs have similar values. The comparator has two outputs: an inverted output used to monitor the state of the circuit (pulse out) and a direct output which goes low when an avalanche is detected, and which stays low while the quenching pulse is applied to the APD. The negative edge of the comparator output clocks a flip-flop, $U4$, applying "ON" voltages to the switch, $U5$, which applies the quenching pulse. The comparator output, delayed by a fixed 20-nsec delay line, $U3$, presets the flip-flop, which turns off the switch and ends the quenching pulse. The Zener diodes, $Z2$ and $Z3$, are used to offset the logic level, as required for the switch control signals. When the quenching pulse ends, steady state is resumed, the comparator output goes high, and the circuit is ready to detect another avalanche.

III. Test Results

The active quenching circuit was tested in stages in order to effectively assess its performance separately from the effects of the avalanche diode. Results indicate that the temporal characteristics of the diode are different from the assumptions used in the quenching circuit design. First, the avalanche current does not increase quickly and smoothly. It takes a few nanoseconds rather than subnanoseconds to build up, and the current may not build monotonically. This was deduced by observing the circuit's output characteristics since the diode current could not be observed directly without affecting its response. In addition, for the Geiger mode of operation, a 3–5-V quenching pulse appears to be insufficient for terminating the triggered avalanche current. The above results indicate that new requirements are needed for a future active quenching circuit design. The following is a more detailed description of the tests and measurements conducted on the active quenching circuit.

A. Initial Circuit Characterization

The active quenching circuit was first assembled in an open-loop configuration to test the signal-propagation delay and rise time. $U2$ and $U4$ were disconnected from each other, and $D1$ was left out. $U4$ was driven by an external clock from a signal generator. The delay and the pulse rise time at the output of the switch were measured with the following results: (1) The delay from comparator output to the switch output was 2.4 nsec and (2) the quenching pulse rise time (to 90 percent of peak value) measured without $D1$ was 6 nsec, and with $D1$, it was 10 nsec. The rise time with $D1$ is longer because the comparator input must return to within a few tens of millivolts of its steady-state value. The capacitive reactance of $D1$ effectively slows the return of the comparator input to its nominal value.

Next, the loop was closed (again without $D1$), and a small signal square wave was superimposed on VN to simulate avalanche events. When the circuit proved to work as a closed loop, i.e., the pulses superimposed at VN caused quenching pulses to be generated at $U5$, it was connected properly, as shown in Fig. 3, with $D1$ included and activated. Output pulse widths measured at the inverted output of $U2$ were quite long—100 nsec for a fixed delay line of 20 nsec (see Fig. 4). That is, the circuit took 80 nsec to respond to the reset signal, an order of magnitude slower than with the APD absent.

As can be seen in Fig. 4, there was a delay of approximately 80 nsec from the rising edge of the quenching pulse (upper trace) to the falling edge of pulse out (lower trace). This was the circuit reset time with the APD present.

When the oscilloscope was triggered on the falling edge of pulse out, a quenching pulse immediately preceding it was observed. Figure 5 shows pulse out, the lower trace, and the quenching pulse at the control input of $U5$, the upper trace. This suggests that a second, later quenching pulse is required to actually reset $D1$, although the occurrence of this later quenching pulse would be the result of noise.

B. Test Configuration 1

In order to characterize the circuit response time, the quenching circuit was connected as shown in Fig. 6. A signal was injected into a spare comparator, using it to generate fast-rising square waves to clock the flip-flop. The flip-flop generated quenching pulses that caused the circuit comparator to change output states. The APD was

connected to the quenching node, but it was biased below breakdown so that no self-sustaining avalanches occurred. This test enabled the authors to measure the time required for the comparator to respond to the drop in current (simulating an avalanche event), as well as the time required to restore the circuit to its quiescent state following the end of the quenching pulse.

Figure 7 shows the comparator's negative input (top) and its output, pulse out (bottom). Since the negative input falls gradually, adjusting the reference voltage (at the input of $U2$) will time-shift the inputs' crossover point and, hence, affect the width of the output pulse. By raising this threshold voltage, the comparator's output pulse width can be decreased to 10 nsec. The increase in the threshold would mean that the diode current will rise to a higher value before an avalanche is detected. If the threshold is increased, the circuit response time is but a few nanoseconds (even with the APD attached and biased below breakdown). Yet, in normal overbiased operation, pulse widths are much longer. This implies that the APD is not capable of being quenched within the quenching pulse width possibly because the quenching pulse is not of sufficient amplitude to take the APD far enough below breakdown to quench an avalanche. Unfortunately, the use of the GaAs circuitry, which enabled the generation of short propagation-delay signals, restricts pulse amplitudes to 4–5 V.

Figure 8 shows the quenching pulse when the switch is driving only the 500-ohm oscilloscope probe. Under this condition, the switch is able to provide a clean, sharp transition. In the circuit, however, the switch is driving a capacitive load (the APD). The top trace of Fig. 9 shows that the quenching pulse when driving the APD is neither sharp, deep, nor smooth, which implies that the APD may not be quenched by any given quenching pulse. The bottom trace of Fig. 9 records the pulse out produced by the quenching circuit when driving the APD.

C. Test Configuration 2

Test Configuration 2 was a modification of the original circuit, designed to clarify the following problem observed in Test Configuration 1: The GaAs comparator requires a subnanosecond monotonically rising edge, or it will oscillate. The comparator responded to subnanosecond test signals without oscillations. Yet, when the APD avalanched, the observed oscillations suggest that the APD's current increase is relatively slow and possibly jittery as well. The recharging of the APD plus the comparator response time contributed about 12 nsec to the total circuit reset time as observed at pulse out. The GaAs single-pole single-throw switch, a Gigabit Logic S0030, has 0.5-nsec switching time for a 4.5-V pulse when properly terminated with 50 ohms. In the original circuit, however, the switch is terminated in the quenching node, which consists of the APD, another diode, and a resistor. This impedance mismatch slows down the pulse response.

Test Configuration 2 is shown in Fig. 10. The switch, the transimpedance amplifier, and the Schottky barrier diode were eliminated. The quenching pulse was reduced to an amplitude of 2 V, by using proper termination. There was less amplification of avalanche current so the comparator operated close to the threshold.

As can be seen from Fig. 11, pulses of about 40-nsec duration can be obtained on pulse out (lower trace). The 20-nsec quenching pulse (upper trace) appears square at the flip-flop output. However, the comparator shows 500-MHz oscillations around the rising edge.

IV. Summary

A high-speed circuit to quench avalanche breakdowns resulting from photons incident upon an avalanche photodiode has been designed, fabricated, and tested. The active quenching circuit takes 2.4 nsec following the detection of an avalanche breakdown of the Geiger-mode APD to initiate a quenching pulse. Approximately 12 nsec are required for the APD to recharge and for the comparator to reset following termination of a quenching pulse. In order to minimize response time, the circuit was designed for a maximum quenching pulse amplitude of 4.5 V, which proved insufficient for reliably quenching actual avalanches. Pulse amplitudes of 10–25 V have been used by other groups [7]. Testing showed that the active quenching circuit would eventually quench and reset the over-biased APD after approximately 80 nsec, but that the quenching probably resulted from noise. Future development of active quenching circuits will involve different requirements, such as larger quenching pulses at a matched impedance node and a subnanosecond, highly sensitive comparator.

Acknowledgments

The authors would like to thank Dr. Chien-Chung Chen for his preparatory work and many helpful discussions, Long Tuyen Ly for excellent craftsmanship in assembling the active quenching circuit, and the application engineers of Gigabit Logic.

References

[1] R. J. McIntyre, "On the Avalanche Initiation Probability of Avalanche Diodes Above the Breakdown Voltage," *IEEE Transactions on Electron Devices*, vol. ED-20, no. 7, pp. 637–641, July 1973.

[2] R. J. McIntyre, "Recent Developments in Silicon Avalanche Photodiodes," *Measurement*, vol. 3, no. 4, pp. 146–152, October–December 1985.

[3] D. L. Robinson and B. D. Metscher, "A Cooled Avalanche Photodiode With High Photon Detection Probability," *TDA Progress Report 42-87*, vol. July–September 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 41–47, November 15, 1986.

[4] D. L. Robinson and D. A. Hays, "Photon Detection With Cooled Avalanche Photodiodes: Theory and Preliminary Experimental Results," *TDA Progress Report 42-81*, vol. January–March 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 9–17, May 15, 1985.

[5] D. L. Robinson and B. D. Metscher, "Photon Detection With Cooled Avalanche Photodiodes," *Applied Physics Letters*, vol. 51, no. 19, pp. 1493–1494, November 9, 1987.

[6] C.-C. Chen, "Effect of Detector Dead Time on the Performance of Optical Direct-Detection Communication Links," *TDA Progress Report 42-93*, vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 146–154, May 15, 1988.

[7] A. W. Lightstone and R. J. McIntyre, "Photon Counting Silicon Avalanche Photodiodes for Photon Correlation Spectroscopy," paper presented at the O.S.A. topical meeting on photon correlation techniques and applications, Washington, D.C., May 30–June 2, 1988.

[8] R. G. W. Brown, R. Jones, J. G. Rarity, and K. D. Ridley, "Characterization of Silicon Avalanche Photodiodes for Photon Correlation Measurements. 2: Active Quenching," *Applied Optics*, vol. 26, no. 12, pp. 2383–2389, June 15, 1987.

[9] S. Cova, A. Longoni, and G. Ripamonti, "Active-Quenching and Gating Circuits for Single-Photon Avalanche Diodes (SPADs)," *IEEE Transactions on Nuclear Science*, vol. NS-29, no. 1, pp. 599–601, February 1982.

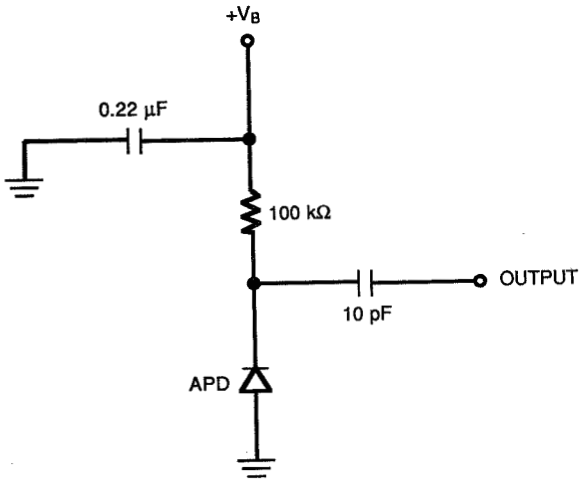


Fig. 1. Passive quenching circuit. Dead time of the avalanche photodiodes using this circuit is 6 μsec.

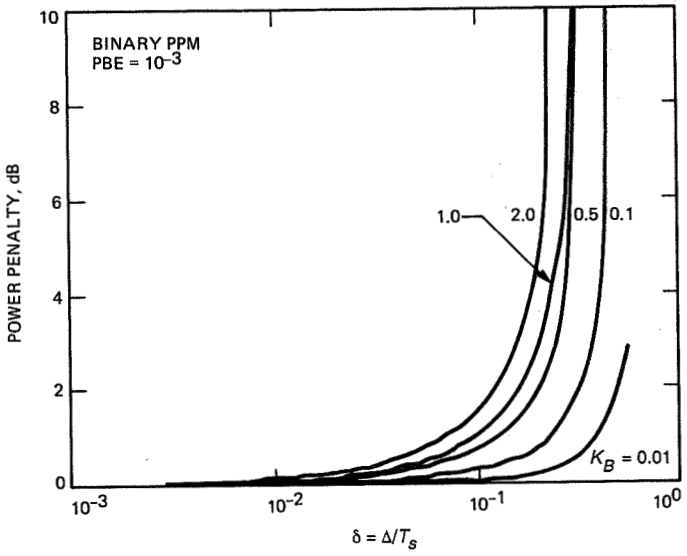


Fig. 2. Power penalty due to detector dead time for a binary pulse position modulation receiver at different background intensities.

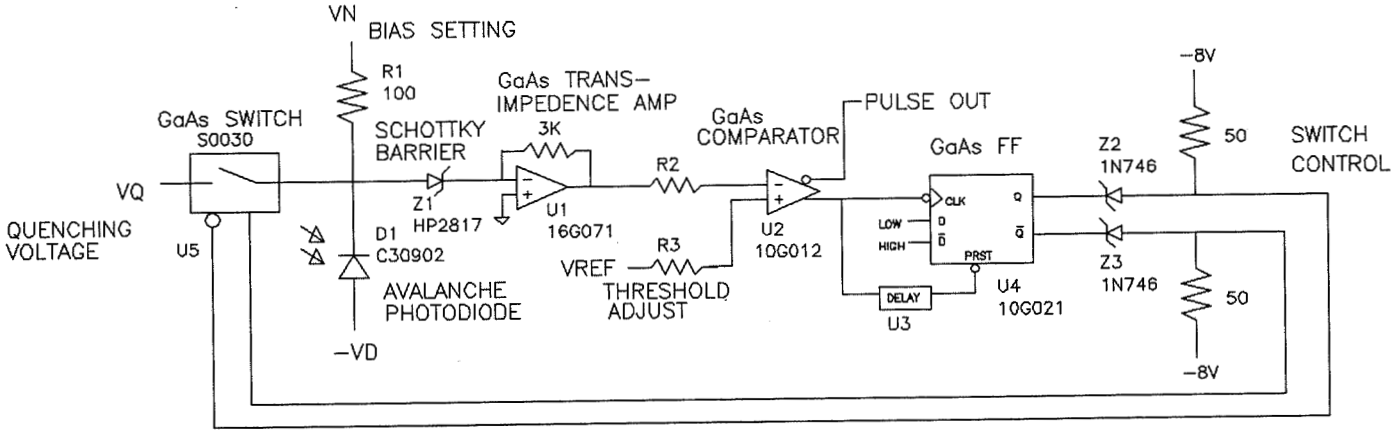


Fig. 3. Active quenching circuit for avalanche photodiodes using GaAs technology.

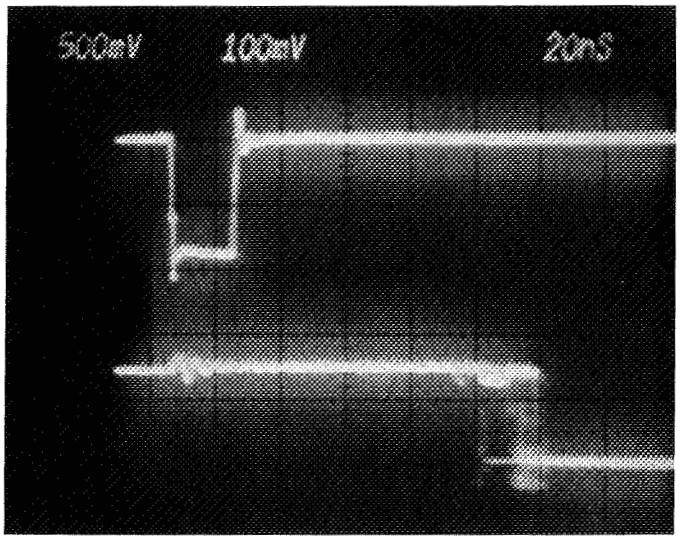


Fig. 4. Triggering the scope on the rising edge of the quenching pulse control signal (upper trace) shows the delay to pulse out reset (lower trace) is 80 nsec.

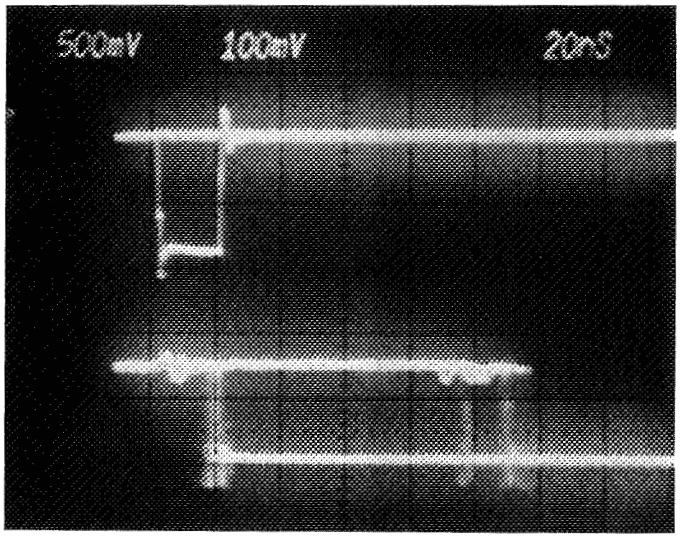


Fig. 5. Triggering the scope on the falling edge of pulse out (bottom trace) shows that a quenching pulse preceded it (upper trace).

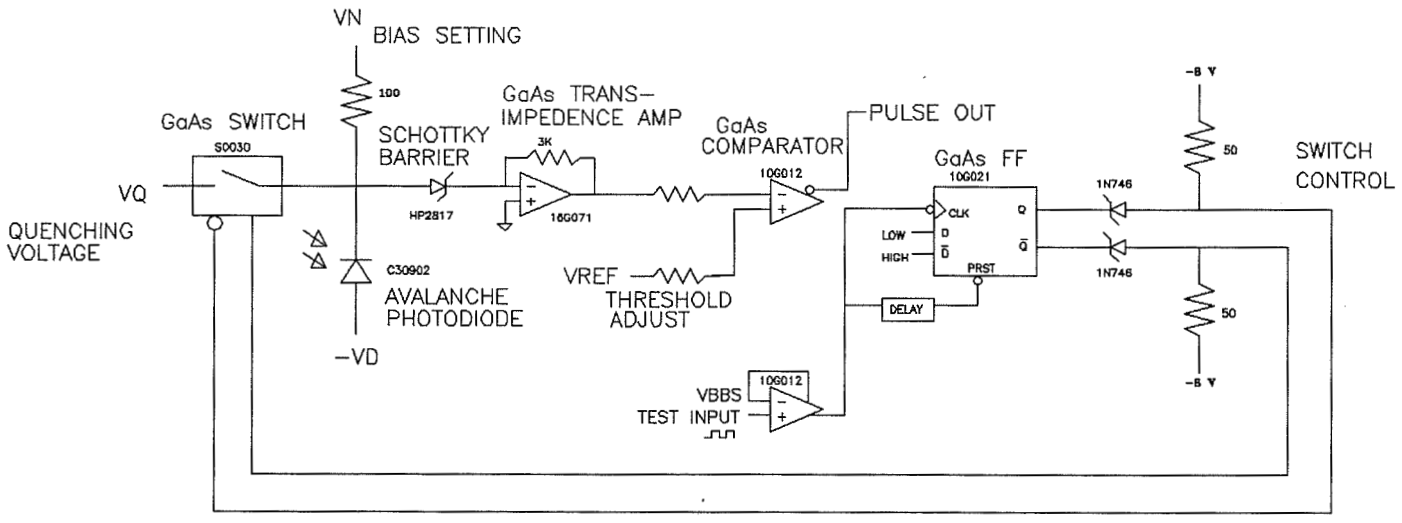


Fig. 6. Test configuration: quenching pulses occur, but no avalanche.

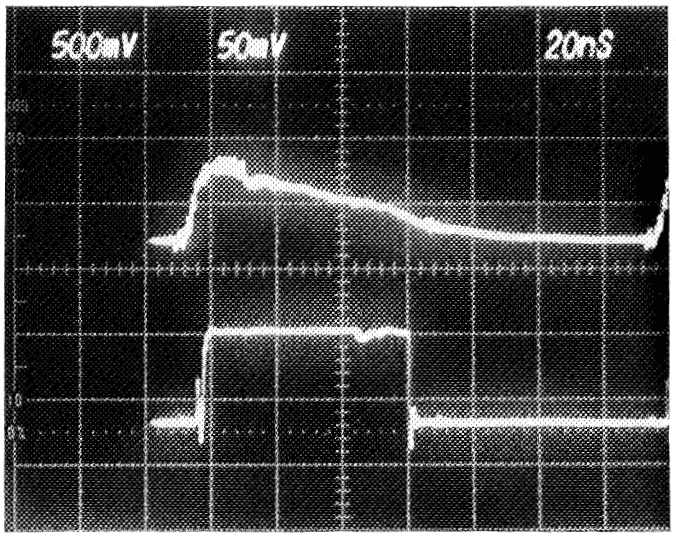


Fig. 7. Comparator input (top trace) versus output (bottom trace) shows that pulse width depends on the threshold.

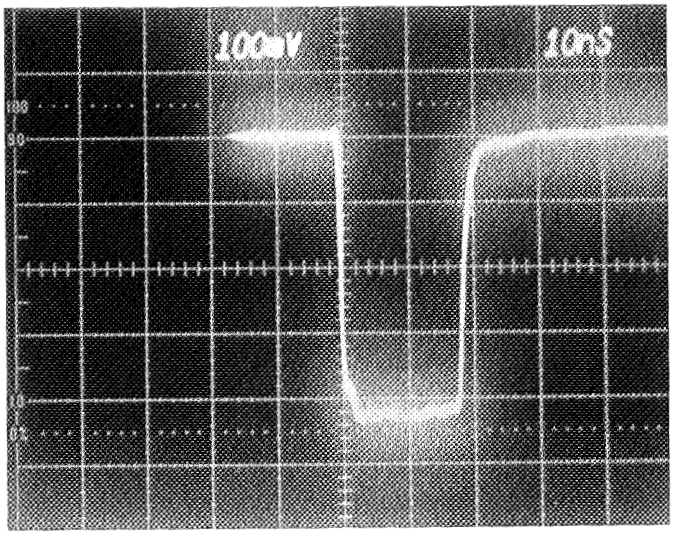


Fig. 8. Quenching pulse, *RL* = 500 ohms.

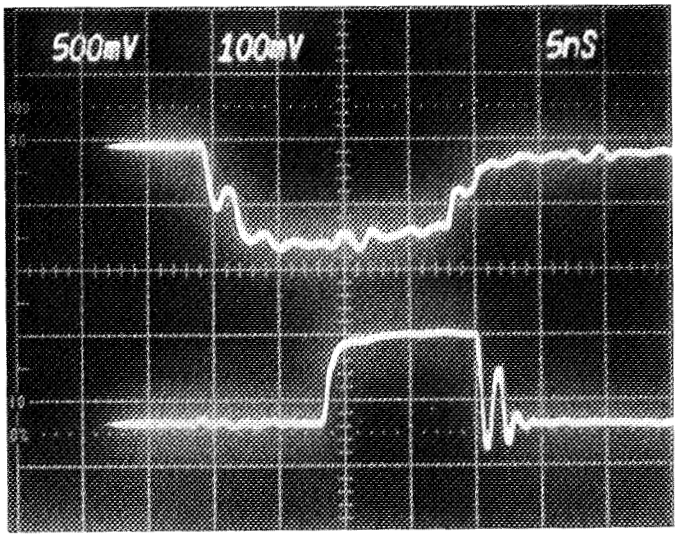


Fig. 9. Top trace: quenching pulse driving the APD; bottom trace: pulse out.

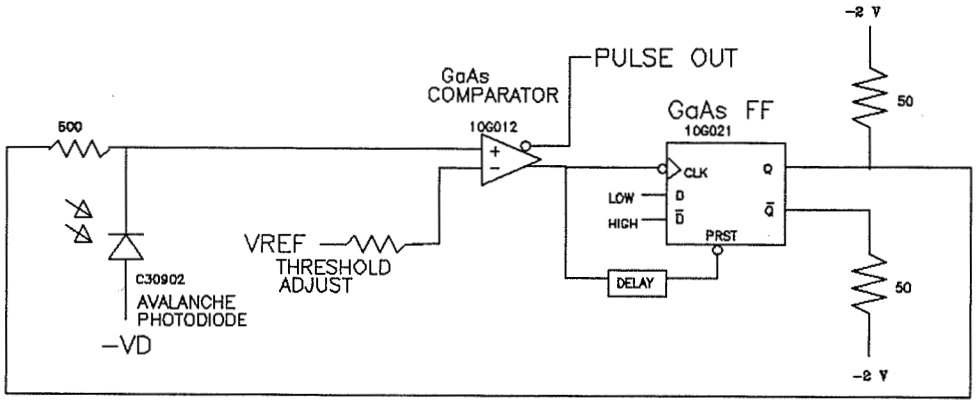


Fig. 10. Test configuration: eliminating switch, transimpedance amplifier, and Schottky diode.

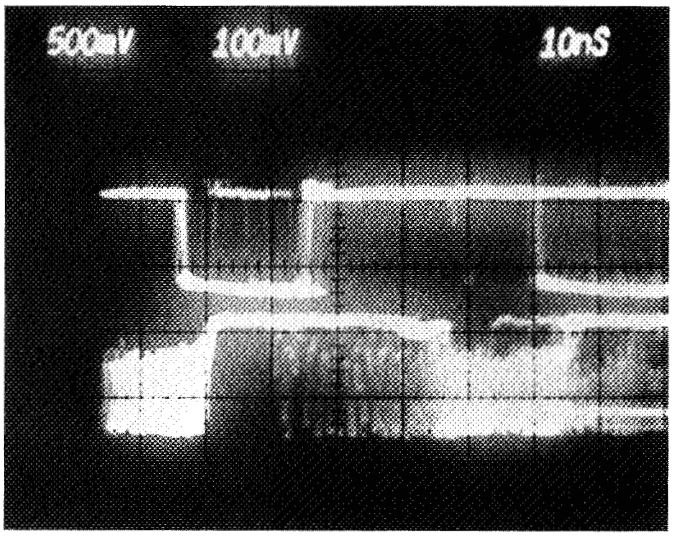


Fig. 11. Top trace: quenching pulse, 20 nsec; bottom trace: pulse out, 40 nsec.

Analysis of a Multiple Reception Model for Processing Images From the Solid-State Imaging Camera

T.-Y. Yan
Communications Systems Research Section

This article describes a detection model to identify the presence of Galileo optical communications from an Earth-based Transmitter (GOPEX) signal by processing multiple signal receptions extracted from the camera images. The model decomposes a multi-signal reception camera image into a set of images so that the location of the pixel being illuminated is known a priori and the laser can illuminate only one pixel at each reception instance. Numerical results show that if effects on the pointing error due to atmospheric refraction can be controlled to between 20–30 μrad, the beam divergence of the GOPEX laser should be adjusted to be between 30–40 μrad when the spacecraft is 30,000,000 km away from Earth. Furthermore, increasing beyond five the number of receptions for processing will not produce a significant detection probability advantage.

I. Introduction

The trajectory of the Galileo spacecraft after the second Earth flyby in December 1992 allows Galileo optical communications from an Earth-based Transmitter (GOPEX) demonstration against a dark-Earth background [1]. Furthermore, as the spacecraft rises to an elevation as high as 55 deg, the deleterious effects of refraction associated with the laser transmissions through the atmosphere will be reduced substantially. Because of these two factors, the Earth-2 flyby has always been regarded as a prime opportunity for the GOPEX demonstration.

Against a dark-Earth background, the GOPEX laser strength received at the spacecraft will produce adequate signal to noise levels, even when the spacecraft is 30,000,000 km away. However, pointing errors will significantly degrade the signal strength at the spacecraft and subsequently reduce the probability of detecting a laser pulse by the solid-state-imaging (SSI) camera. Detection of the laser signal is of particular concern because, as the spacecraft recedes from Earth, the signal strength at the camera decreases. Different strategies for processing the images must therefore be used to ascertain whether the laser signal was detected by the camera.

The strategy for laser transmissions to Galileo will differ over the proposed 28-day course of experiments. Close to Earth, the camera shutter will remain open for only a

few tens of milliseconds to prevent lateral blooming of the camera pixels imaging the bright Earth onto those pixels containing the laser transmissions against a dark-Earth background. At the 30-Hz maximum laser-transmission rate, only one or two transmitted pulses are expected to be captured in a single camera frame. However, as the spacecraft recedes, the Earth will subtend fewer pixels in the focal plane of the SSI, and the camera shutter could be kept open for a longer period (approximately 400 msec) before lateral blooming affects the laser-illuminated pixels. Figure 1 shows a sample image of the GOPEX laser transmission.

It is expected that for a given set of input conditions, the probability of detecting the laser transmission increases with the number of received laser pulses available for processing in a given frame. Thus, at the longer ranges, where the laser signal is reduced, scanning the camera across the image of the Earth to illuminate several pixels with the laser transmission will serve not only to augment the deep-space communication demonstration, but also as a tool in processing the returned data to enhance the probability of identifying the transmitted laser signal.

This article describes a mathematical model to analyze the detection probability for the case of several laser pulses on a single frame. This analysis addresses the transmission and detection strategies for the longer ranges when the SSI camera is to be scanned across the image of the Earth. In particular, numerical examples are given for when the spacecraft is 30,000,000 km away from the Earth.

The multiple independent reception analysis presented here extends the single reception-detection model in [1]. It is the second step in the development of a comprehensive model to analyze the probability of detecting the laser transmission under actual conditions expected during the Earth-2 flyby. In the current analysis, the effects of Earth rotation and the relative velocity between the spacecraft and Earth are ignored, and it is assumed that the GOPEX laser pulse illuminates a single pixel with a known location.

II. Mathematical Model

The analysis begins with the hypothesis that when the GOPEX laser pulse is transmitted, it is called H_1, and when the laser pulse is not transmitted, it is called H_0. A total of N receptions are assumed to be available for processing. The detection processor will not determine which hypothesis is true until all N receptions are processed. Further, the receptions are assumed to be sufficiently apart so that they can be considered to be independent of one another. Let r_i be the received number of photoelectrons at the ith reception. Under the two hypotheses, then

$$H_1 \;:\; r_i = s_i + e_i + n_i$$

$$H_0 \;:\; r_i = e_i + n_i, \qquad i = 1, \ldots, N \tag{1}$$

where s_i is a Gaussian random variable representing the number of photoelectrons due to the GOPEX laser with mean μ_s and variance σ_s; e_i is a Gaussian random variable representing the number of received photoelectrons due to earthshine with mean μ_e and variance σ_e; and n is a Gaussian random variable representing other disturbances with zero mean and variance σ_n. The random variables s_i, e_i, and n_i are uncorrelated relative to one another.

Define $\Lambda(\mathbf{r})$ to be the likelihood ratio of the two hypotheses where $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ is the received random vector. Then

$$\Lambda(\mathbf{r}) = \prod_{i=1}^{N} \left\{ \frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(r_i - \mu_s - \mu_e)^2}{2\sigma_1^2}}}{\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(r_i - \mu_e)^2}{2\sigma_0^2}}} \right\} \tag{2}$$

where $\sigma_1^2 = \sigma_s^2 + \sigma_e^2 + \sigma_n^2$ and $\sigma_0^2 = \sigma_e^2 + \sigma_n^2$.

Under the Neyman-Pearson criterion, the probability of detection can be maximized by setting the detection threshold γ to the log likelihood ratio for a given probability of false alarm. From Eq. (2), setting $\ell(\mathbf{r}) = \ln \Lambda(\mathbf{r})$, the optimal detection processor is given by

$$\ell(\mathbf{r}) = -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N} (r_i - \mu_s - \mu_e)^2 + \frac{1}{2\sigma_0^2} \sum_{i=1}^{N} (r_i - \mu_e)^2 + N \ln\left(\frac{\sigma_0}{\sigma_1}\right) \tag{3}$$

The processor will declare the existence of the GOPEX laser pulse if Eq. (3) exceeds the threshold γ. In Eq. (3), the optimal processor is the difference of two quadratic forms. Each of the summations in Eq. (3) has a noncentral χ^2 distribution. Although the probability density function of the random variable ℓ (the log likelihood ratio) can be computed numerically, the approximation provided

by using the moment-generating function is much simpler and intuitive. It is generally adequate to predict the performance of the optimal detector.

Define P_{fa} to be the probability of false alarm (i.e., the laser is transmitted when it is not) and define P_d to be the probability of detection (i.e., the laser is transmitted when it is). From [2], the probability of false alarm and the probability of detection can be approximated by

$$P_{fa} \approx e^{\xi(s)-s\dot{\xi}(s)+s^2/2\ddot{\xi}(s)} \cdot Q\left(s\sqrt{\ddot{\xi}(s)}\right) \qquad (4)$$

and

$$P_d \approx e^{\xi(s)+(1-s)\dot{\xi}(s)+(s-1)^2/2\ddot{\xi}(s)} \cdot Q\left((1-s)\sqrt{\ddot{\xi}(s)}\right) \qquad (5)$$

where $\xi(s)$, $\dot{\xi}(s)$, and $\ddot{\xi}(s)$ are given in Appendix A, and

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$$

The approximations in Eqs. (4) and (5) will give exact solutions for cases of equal variances, i.e., $\sigma_1 = \sigma_0$. This corresponds to the case of deterministic GOPEX laser transmission of μ_s. For example, by setting $\sigma_1 = \sigma_0$ and $N = 1$, this model reduces to the previous model reported in [1]. Since the approximation makes use of the Law of Large numbers, results presented here represent a tight approximation to the exact solutions for $N \geq 3$. In general, for any given probability of false alarm, a corresponding s_f can be computed by using Eq. (A-8). Substituting the resulting s_f in Eq. (A-6) gives the threshold γ. The resulting detection probability is then given by Eq. (5), which uses the same s_f.

III. Numerical Results

This section presents results of the multiple reception-detection analysis. In these calculations, a laser output energy of 0.25J at the 0.532-μm wavelength is assumed. Since the GOPEX demonstration will be conducted against a dark-Earth background, a nominal photoelectron count of 600 photon background noise is assumed when the camera shutter remains open for half a second. This count is approximately 10 percent of the background earthshine used for the Earth-1 analysis.

Figures 2 and 3 show some of the computation results. Without loss of generality, the probability of false alarm is taken to be no more than 1×10^{-6}. The noise variance of the GOPEX laser is extracted from [2]. Figure 2 plots the detection probability as a function of the beam divergence of the GOPEX laser pulse when the spacecraft is 30,000,000 km away from Earth. The curves are parametrized by the number of receptions available for processing. The pointing error is assumed to be 25 μrad for the computation. It can be seen that the maximum probability of detection increases with the number of processed receptions. However, processing more than four receptions will result in only minor improvements of the detection probability. In all four cases, the maximum probability of detection occurs between 25–40 μrad laser-beam divergence.

Figure 3 shows the probability of detection when the spacecraft is again 30,000,000 km away from Earth, with three processing receptions. The plots are parametrized by pointing errors of 20, 25, and 30 μrad. When the beam divergence is small, as compared with the pointing error, the probability of detecting the presence of the GOPEX laser pulse degenerates to the probability of searching for the GOPEX laser pulse within the space spanned by the pointing error. Consequently, the detection probability decreases as the beam divergence is reduced. On the other hand, since the intensity of the GOPEX laser pulse decreases with increasing beam divergence, the detection probability will decrease as the beam diverges. It can be seen that for all three cases, the maximum detection probability is achieved when the beam divergence is between 30–40 μrad.

These figures indicate that if the pointing error of the GOPEX laser pulse is confined to between 20–30 μrad, the beam divergence should be maintained between 30–40 μrad to maximize the detection probability when the spacecraft is 30,000,000 km from Earth. Furthermore, the detection probability can be enhanced by processing multiple receptions within the range of interest.

IV. Summary

This article describes a detection model to identify GOPEX laser transmissions by processing multiple receptions extracted from the camera image. The approach presented here is valid for $N > 3$, even if the distributions of s_i, e_i, and n_i are non-Gaussian. Numerical results show that if effects of the pointing error due to atmospheric refraction can be controlled between 20–30 μrad, the beam divergence of the GOPEX laser should be adjusted to be-

tween 30–40 μrad when the spacecraft is 30,000,000 km from Earth. Increasing the number of receptions for processing beyond four will have diminishing returns.

There are several assumptions made to reach these conclusions. First, the model assumes a priori that the location of the pixel being illuminated is known and that the laser illuminates only one pixel at each reception instance. Effects of any disturbance due to Earth rotation and the jitter of the spacecraft dynamics are not considered. Further effort will be required to include these effects and will be reported later.

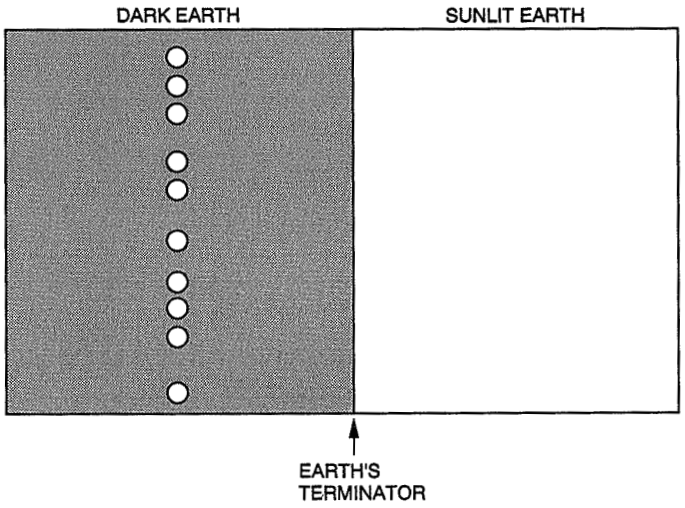


Fig. 1. A sample of GOPEX laser transmission.

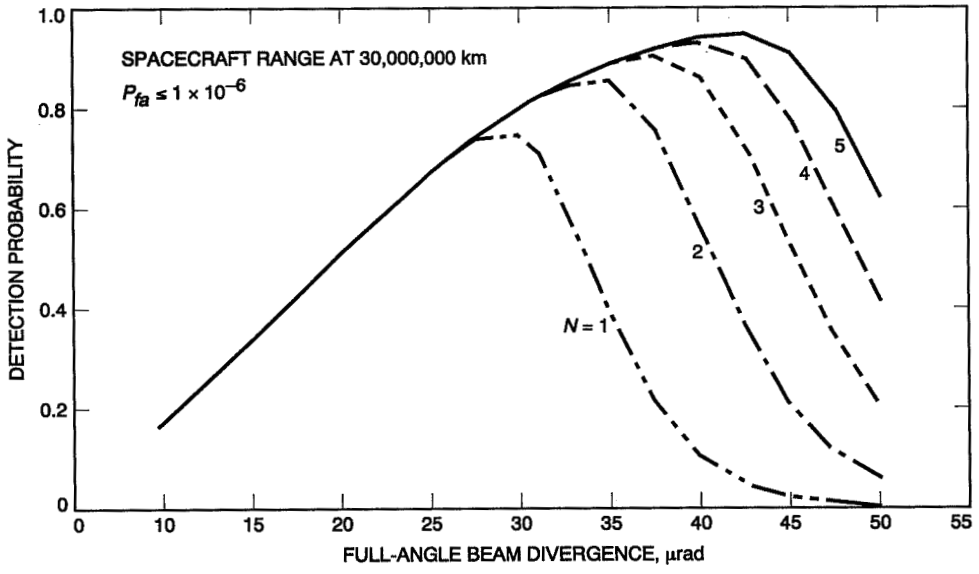


Fig. 2. Effect of increasing number of receptions on the detection probability.

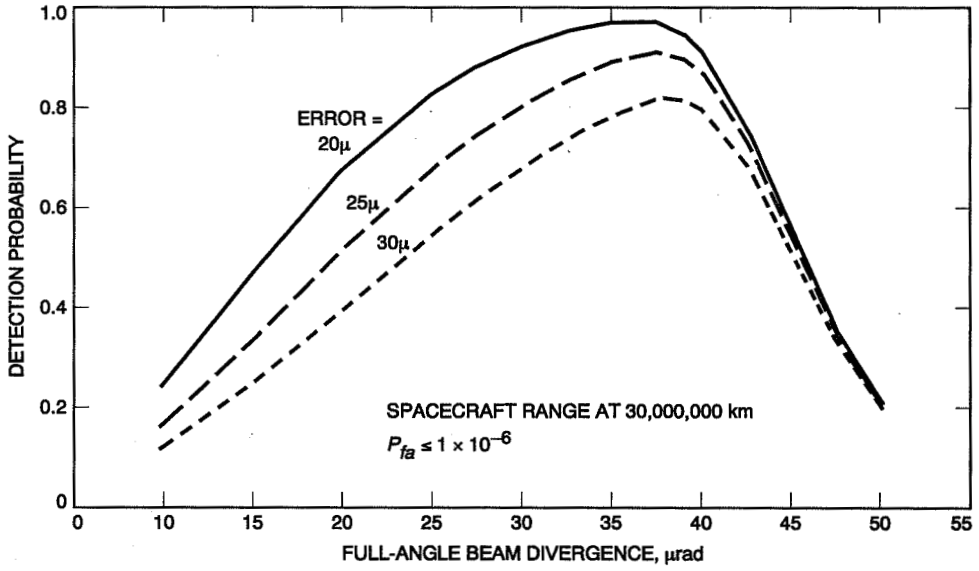


Fig. 3. Degradation of the detection probability due to laser pointing errors.

Appendix A
The Approximation of P_{fa} and P_d

The probability density functions of the received vector $\mathbf{r}$ under the two hypotheses are given respectively by

$$p(\mathbf{r}/H_1) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(r_i-\mu_s-\mu_e)^2/2\sigma_1^2} \tag{A-1}$$

and

$$p(\mathbf{r}/H_0) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(r_i-\mu_e)^2/2\sigma_0^2} \tag{A-2}$$

where $\sigma_1^2 = \sigma_s^2 + \sigma_e^2 + \sigma_n^2$ and $\sigma_0^2 = \sigma_e^2 + \sigma_n^2$.

Let

$$e^{\xi(s)} = \mathrm{E}\left[e^{s\ell(\mathbf{r})}/H_0\right] \tag{A-3}$$

be the moment-generating function of the log likelihood ratio $\ell(\mathbf{r})$ under H_0. Then, substituting Eqs. (A-1) and (A-2) in Eq. (A-3) gives

$$\xi(s) = N \ln \left\{ \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(r_i-\mu_s-\mu_e)^2/2\sigma_1^2} \right]^s \times \left[\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(r_i-\mu_e)^2/2\sigma_0^2} \right]^{1-s} dr_i \right\} \tag{A-4}$$

After some algebraic manipulations, the integral is given by

$$\xi(s) = \frac{N}{2} \left\{ \ln \left(\frac{(\sigma_1^2)^{1-s}(\sigma_0^2)^s}{s\sigma_0^2 + (1-s)\sigma_1^2} \right) + \frac{s(s-1)\mu_s^2}{s\sigma_0^2 + (1-s)\sigma_1^2} \right\} \tag{A-5}$$

Differentiating Eq. (A-5) with respect to s, one has

$$\dot{\xi}(s) = \frac{N}{2} \left\{ \ln\left(\frac{\sigma_0^2}{\sigma_1^2}\right) - \frac{\sigma_0^2 - \sigma_1^2}{s\sigma_0^2 + (1-s)\sigma_1^2} + \mu_s^2 \frac{s^2\sigma_0^2 - (s-1)^2\sigma_1^2}{[s\sigma_0^2 + (1-s)\sigma_1^2]^2} \right\} \tag{A-6}$$

The second derivative of $\xi(s)$ is given by

$$\ddot{\xi}(s) = \frac{N}{2} \left\{ \left(\frac{\sigma_0^2 - \sigma_1^2}{s\sigma_0^2 + (1-s)\sigma_1^2} \right)^2 + 2\mu_s^2 \frac{\sigma_0^2\sigma_1^2}{[s\sigma_0^2 + (1-s)\sigma_1^2]^3} \right\} \tag{A-7}$$

From [3], the probability of false alarm is

$$P_{fa} \approx e^{\xi(s) - s\dot{\xi}(s) + (s^2/2)\ddot{\xi}(s)} \cdot Q\left(s\sqrt{\ddot{\xi}(s)}\right) \tag{A-8}$$

and the probability of detection is given by

$$P_d \approx e^{\xi(s) + (1-s)\dot{\xi}(s) + [(s-1)^2/2]\ddot{\xi}(s)} \cdot Q\left((1-s)\sqrt{\ddot{\xi}(s)}\right) \tag{A-9}$$

The corresponding threshold can be determined as

$$\gamma = \dot{\xi}(s)$$

The function $Q(x)$ is given by

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt$$

It should be noted that $\ddot{\xi}(s)$ must be positive in both Eqs. (A-8) and (A-9) for all values of s. The first term on the right-hand side of Eq. (A-7) is always positive. A sufficient condition for $\ddot{\xi}(s)$ to be positive is that the denominator of the second term remains positive for all values of s. Hence, the variable s is restricted by the following inequality,

$$0 \le s \le \frac{\sigma_1^2}{\sigma_1^2 - \sigma_0^2}$$

References

[1] K. E. Wilson, J. R. Lesh, T.-Y. Yan, J. Schwartz, M. D. Rayman, and S. Wee, "GOPEX: A Deep Space Optical Communications Demonstration With the Galileo Spacecraft," *TDA Progress Report 42-103*, vol. July–September 1990, pp. 262–277, November 15, 1990.

[2] K. Klaasen, M. Clary, and J. Janesick, "Charge-Coupled Device Television Camera for NASA's Galileo Mission to Jupiter," *Optical Engineering*, vol. 23, no. 3, pp. 334–342, May/June 1984.

[3] H. Van Trees, *Detection, Estimation, and Modulation Theory*, New York: John Wiley & Sons, 1968.

The Advanced Receiver II Telemetry Test Results at Goldstone

R. Sadr, R. Bevan, and S. Hinedi
Communications Systems Research Section

This article describes the telemetry performance of the Advanced Receiver II (ARX-II) in tracking and demodulating signals from various deep space missions. The tests were performed at Goldstone, and the spacecraft tracked were Pioneer 10, Pioneer 11, Voyager 2, Magellan, and the International Cometary Explorer (ICE). These missions present a broad range of operating conditions to test the ARX-II in terms of signal-to-noise ratio (SNR) and the symbol rate of the received signal.

I. Introduction

The Advanced Receiver II (ARX-II) is an all-digital receiver capable of performing various tasks [1]. Its Doppler extractor has been tested at Goldstone [2], and its telemetry performance was measured in Compatibility Test Area 21 (CTA21) [3].

In late November 1990, the received signal from Pioneer 11 exhibited unpredictable behavior. The ARX-II was sent to Goldstone to try to acquire and lock onto the received signal and demodulate the symbols. There was strong evidence that the onboard oscillator of the spacecraft was malfunctioning and that a one-way link would not be possible. Data are presented for both the one-way mode, in which the ARX-II was not able to maintain lock and did not demodulate the symbols, and the three-way mode, in which the symbols were detected. While at Goldstone, other spacecraft were tracked, including Voyager 2, Pioneer 10, Magellan, and International Cometary Explorer (ICE). The ARX-II logged data from all these signals, and some segments are presented in this article. In Section II, the ARX-II configuration in SPC-10 and the various spacecraft signals are discussed. The measured parameters are presented and discussed in Section III, followed by the conclusion in Section IV.

II. ARX-II Configuration and Spacecraft Parameters

The ARX-II was placed in SPC-10 and connected to the Very Long Baseline Interferometry (VLBI) downconverter (D/C), as shown in Fig. 1. The output of the VLBI D/C provided an intermediate frequency (IF) signal in the neighborhood of 300 MHz. The signal was further downconverted by the ARX-II, digitized, and processed to obtain carrier lock and symbol detection. In a parallel effort, the received signal was also processed by the Block IV (Blk-IV) receiver/Baseband Assembly (BBA) combination or by the Block III (Blk-III) receiver/Subcarrier Demodulation Assembly/Symbol Synchronization Assembly

(SDA/SSA) combination. Depending on the spacecraft, either the Blk-IV/BBA or the Blk-III/SDA/SSA combination was used by the station operators. Due to the urgency of the assignment, no arrangements were made to obtain data from the station receivers, and as a result, no performance comparison of the various demodulators is attempted in this article. A complete set of tests was performed in CTA21 for performance comparison; [3] contains the results.

As mentioned earlier, several spacecraft were tracked with various signal-to-noise ratios (SNRs) and symbol rates. Table 1 summarizes the key signal parameters from each spacecraft, such as received carrier power-to-noise ratio (P_c/N_0), subcarrier frequency (F_{sub}), symbol rate ($R_{sym} = 1/T_s$ where T_s denotes the symbol period), and symbol signal-to-noise-ratio (SSNR). The parameters shown are the predicts and are typically close to their measured values. Pioneer 10 represents the weakest received signal, with $P_c/N_0 = 12$ dB-Hz, a symbol rate of 32 symbols per second (sps), and SSNR = 1.0 dB. On the other hand, Magellan transmits the signal with the highest data rate, at 537.6 ksps, and with SSNR = 1.5 dB. Actually, both channels of Magellan were tracked with the lowest data rate at 80 sps and a 17.0-dB SSNR. Voyager 2 and ICE signals represent intermediate symbol rates at 1200 sps and 256 sps, respectively.

III. Measured Performance

In all the figures to follow, the parameter sets shown are the ones that the ARX-II used for tracking the received signal from the spacecraft: the loop type indicating the order of the carrier digital phase-lock-loop filter, the bandwidth of the loop B, the update period of the loop T_u, the estimation period of the logged data points T_e, and the Doppler period T_d. In the figures, the steady-state phase error ϕ_{ss}, the symbol SNR, and the frequency deviation for each case are shown.

Pioneer 11 data are presented first, for both one-way and three-way links. In the one-way mode, an attempt was made to lock onto the received carrier on DOY 293. The numerically controlled oscillator (NCO) of the ARX-II was swept at the predicted Doppler rate F_d of 0.2 Hz/sec. The estimated P_c/N_0 and the corresponding steady-state phase error ϕ_{ss} are shown in Figs. 2(a) and 2(b), respectively. Loss of carrier lock was frequently encountered, which was also experienced by the monitor and control at SCP-10. When in lock, the P_c/N_0 was estimated to be about 16 dB-Hz with a 10-deg steady-state phase error. The period of dropping lock was roughly about once every 10 sec. Open-loop data were also obtained using the ARX-II. In this case, the frequency was swept at 0.2 Hz/sec, and a fast Fourier transform (FFT), size = 512, rate = 200 Hz, was performed using 10-min data segments. For an ideal tone with a fixed frequency, one expects to observe a peak at the frequency offset. In Fig. 2(c), the peak is depicted versus time. The mean of the frequency offset was about 18 Hz, with about 40-Hz root-mean-squared (RMS) frequency excursions. Figure 2(d) depicts a histogram of the offset frequency, which is a crude estimate of the probability density function of the offset frequency. From these tests, it was concluded that the signal is experiencing frequency changes on the spacecraft transponder, with a frequency deviation of about 40 Hz.

In Fig. 3, the estimated P_c/N_0 (in dB-Hz) is shown in the three-way mode for Pioneer 11. In this case, the ARX-II was able to maintain lock and demodulate the symbols. The corresponding symbol SNR, steady-state phase error, and received frequency estimates are shown in Fig. 4. The receiver was employing a type-III loop with a 1-Hz bandwidth. Also shown in Fig. 3 are the estimates of the carrier power-to-spectral level for the other spacecraft, namely Voyager 2, Pioneer 10, ICE, and both data channels of Magellan. Pioneer 10 has the lowest received P_c/N_0 at 11 dB-Hz, and Magellan has the highest at about 43 dB-Hz (low data rate channel).

The Magellan data are presented in Figs. 5 and 6, with the low data channel in Fig. 5 and the high data channel in Fig. 6. In both cases, a type-III loop was used, resulting in zero steady-state phase errors, with a carrier loop bandwidth around 25 Hz.

Signals with medium data rates were tracked, namely those of Voyager 2 and ICE. The data are presented in Figs. 7 and 8, respectively. The estimated symbol SNR from ICE was about 1 dB, and from Voyager 2 it was about 2.5 dB. Pioneer 10 was also tracked with roughly a 0-dB symbol SNR. It has the lowest symbol rate at 32 sps, and the estimated parameters are shown in Fig. 9. A type-II loop was used resulting in a nonzero steady-state phase error.

IV. Conclusion

The ARX-II successfully tracked the signals and demodulated the symbols from Pioneer 10, Pioneer 11, Voyager 2, ICE, and Magellan. The overall carrier-to-noise spectral level and symbol SNR estimates in each case were consistent with the predicted values.

Acknowledgments

The authors appreciate the assistance of Remi LaBelle and Miguel Marina of the Radio Frequency and Microwave Subsystems Section in performing these tests at SPC-10, and express thanks to the staff of the TDA Mission Support and DSN Operations Section at SPC-10 for providing excellent and expedient logistical support throughout the duration of these tests.

References

[1] S. Hinedi, "A Functional Description of the Advanced Receiver," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 131–149, February 15, 1990.

[2] S. Hinedi, R. Bevan, H. Del Castillo, P. Kinman, D. Chong, and R. Labelle, "Digital Doppler Extraction Demonstration With the Advanced Receiver," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 160–173, February 15, 1990.

[3] S. Hinedi, R. Bevan, and M. Marina, "The Advanced Receiver II Telemetry Test Results in CTA-21," *TDA Progress Report 42-104*, vol. October–December 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 140–156, January 15, 1991.

Table 1. ARX-II and spacecraft parameters

Spacecraft	DSS	P_c/N_0, dB-Hz	F_{sub}, kHz	R_{sym}, sps	SSNR, dB
ICE	14	20	1.024	256	1.0
Voyager 2	15	26	369	1200	2.5
Magellan					
Low rate	15	33	960	537,600	1.5
High rate	15	43	22.5	80	17.0
Pioneer 10	14	12	32.768	32	1.0
Pioneer 11	14	17	32.768	128	2.0

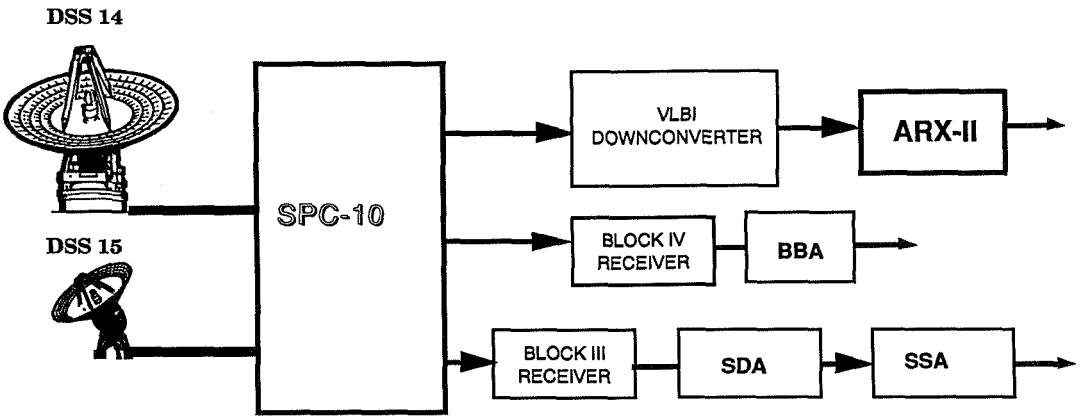


Fig. 1. SPC-10 test configuration with ARX-II.

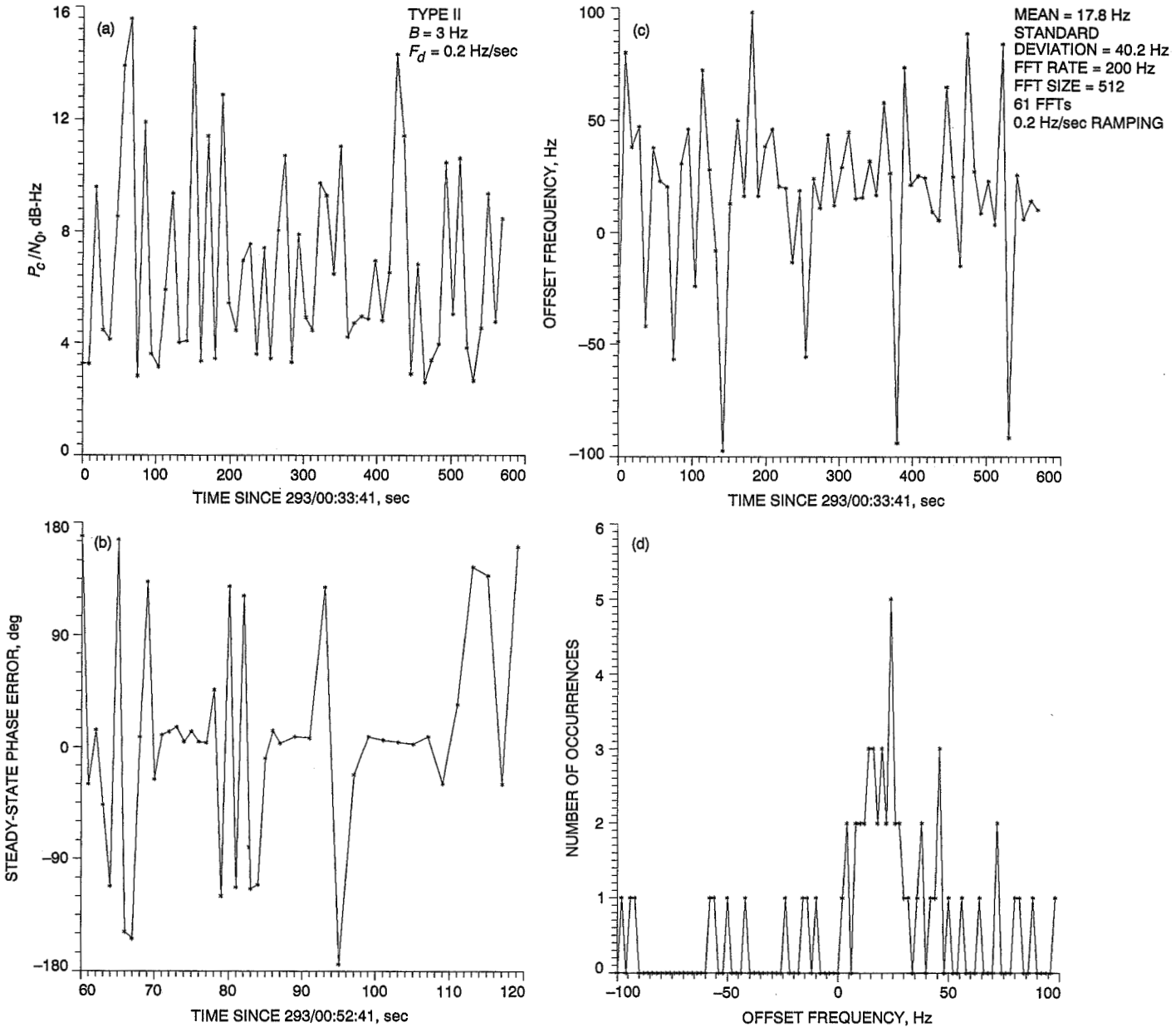


Fig. 2. Pioneer 11 one-way test results on DOY 293: (a) P_c/N_0; (b) ϕ_{ss}; (c) ΔF_c; and (d) histogram of ΔF_c.

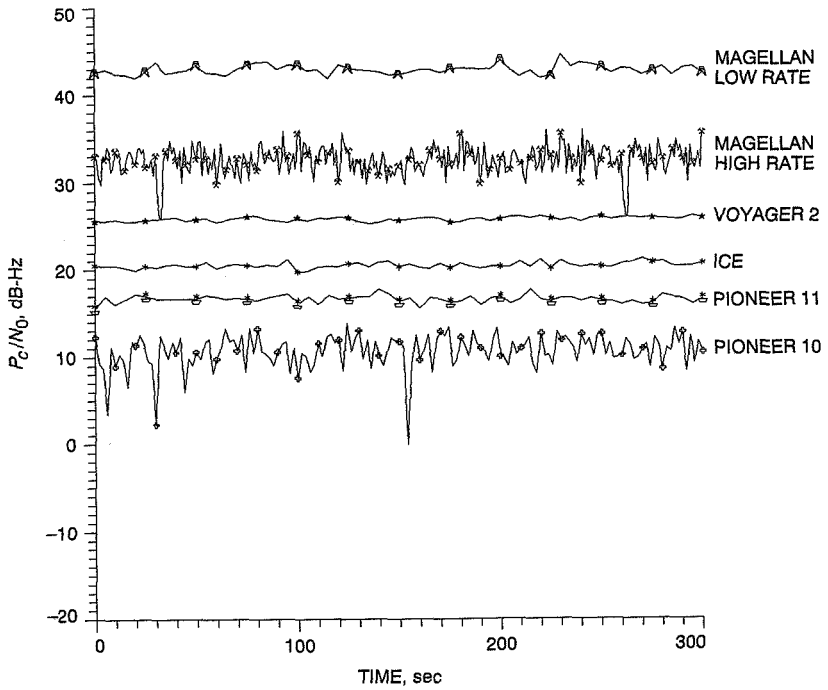


Fig. 3. Summary of test results for carrier-to-noise power (P_c/N_0 for various spacecraft; DOY 293, 299).

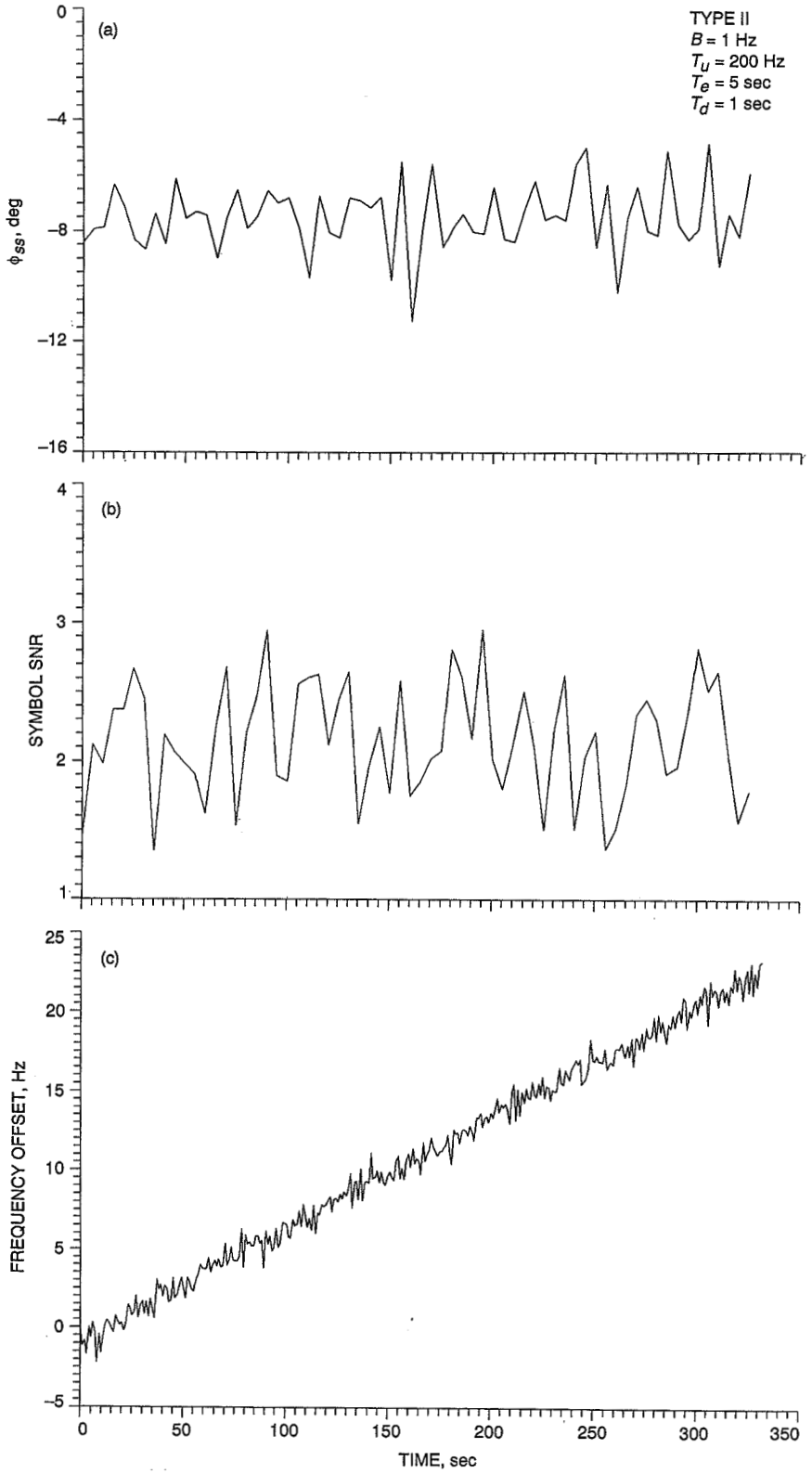


Fig. 4. Pioneer 11 test results on DOY 299: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

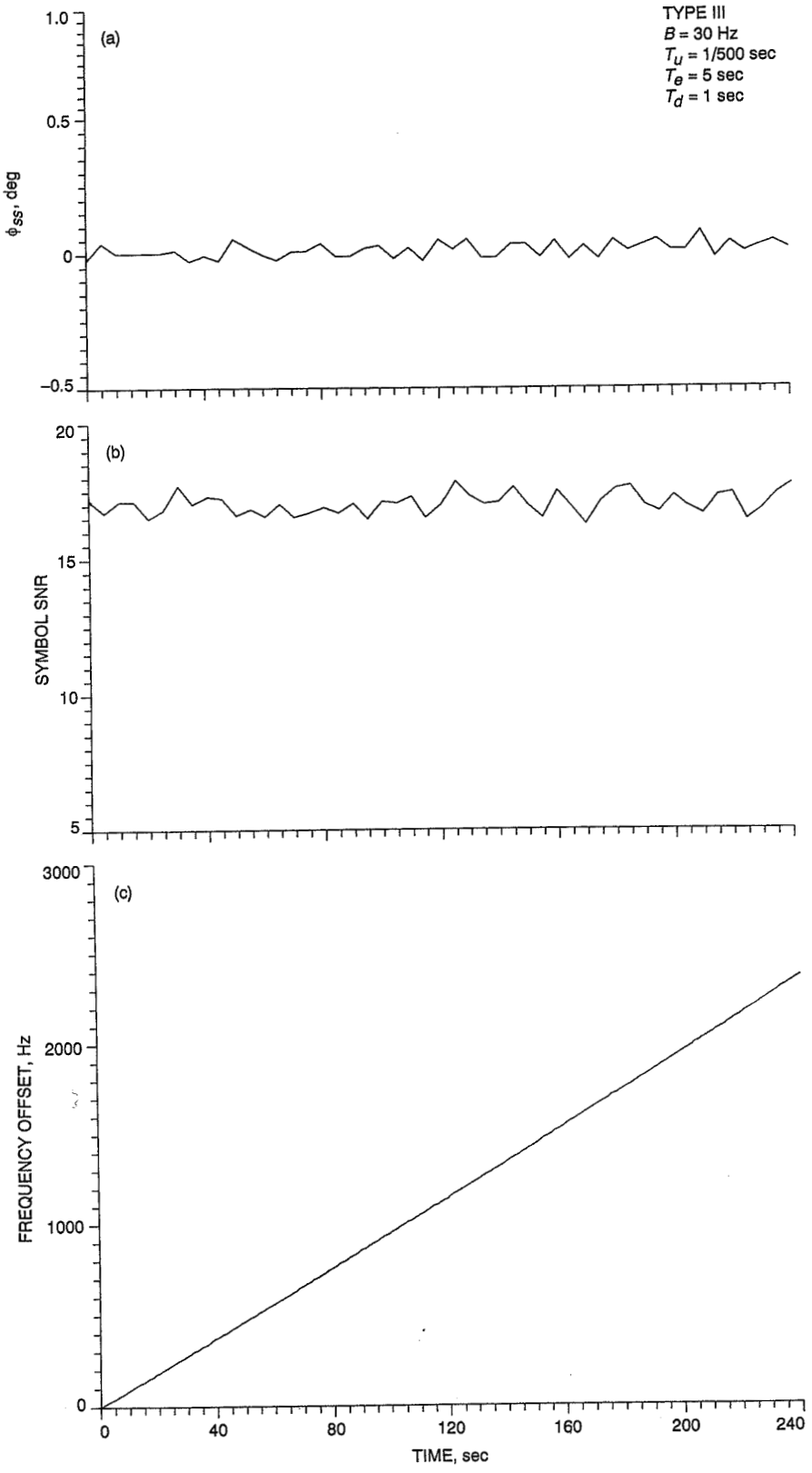


Fig. 5. Magellan low data rate test results on DOY 293: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

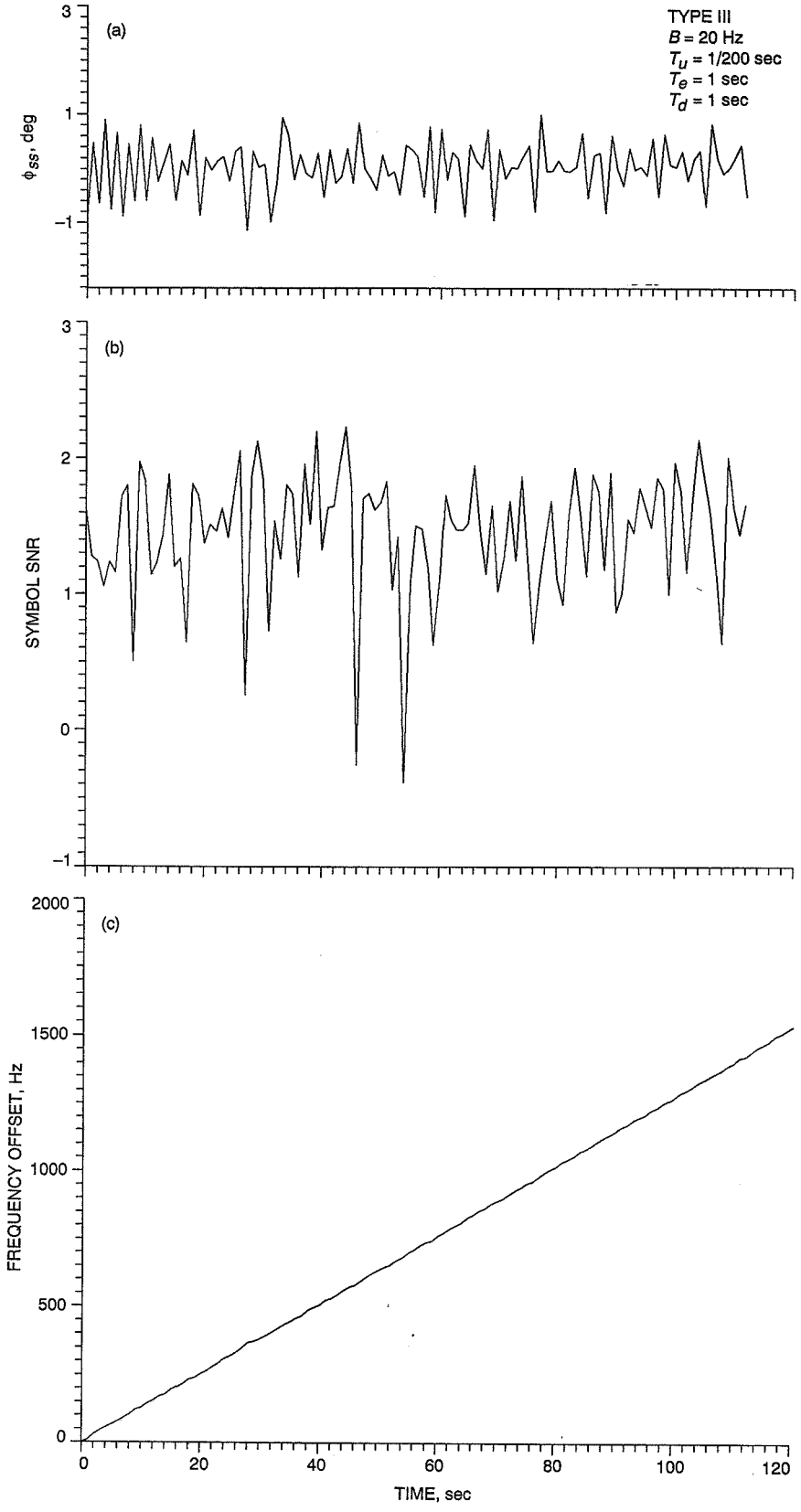


Fig. 6. Magellan high data rate test results on DOY 293: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

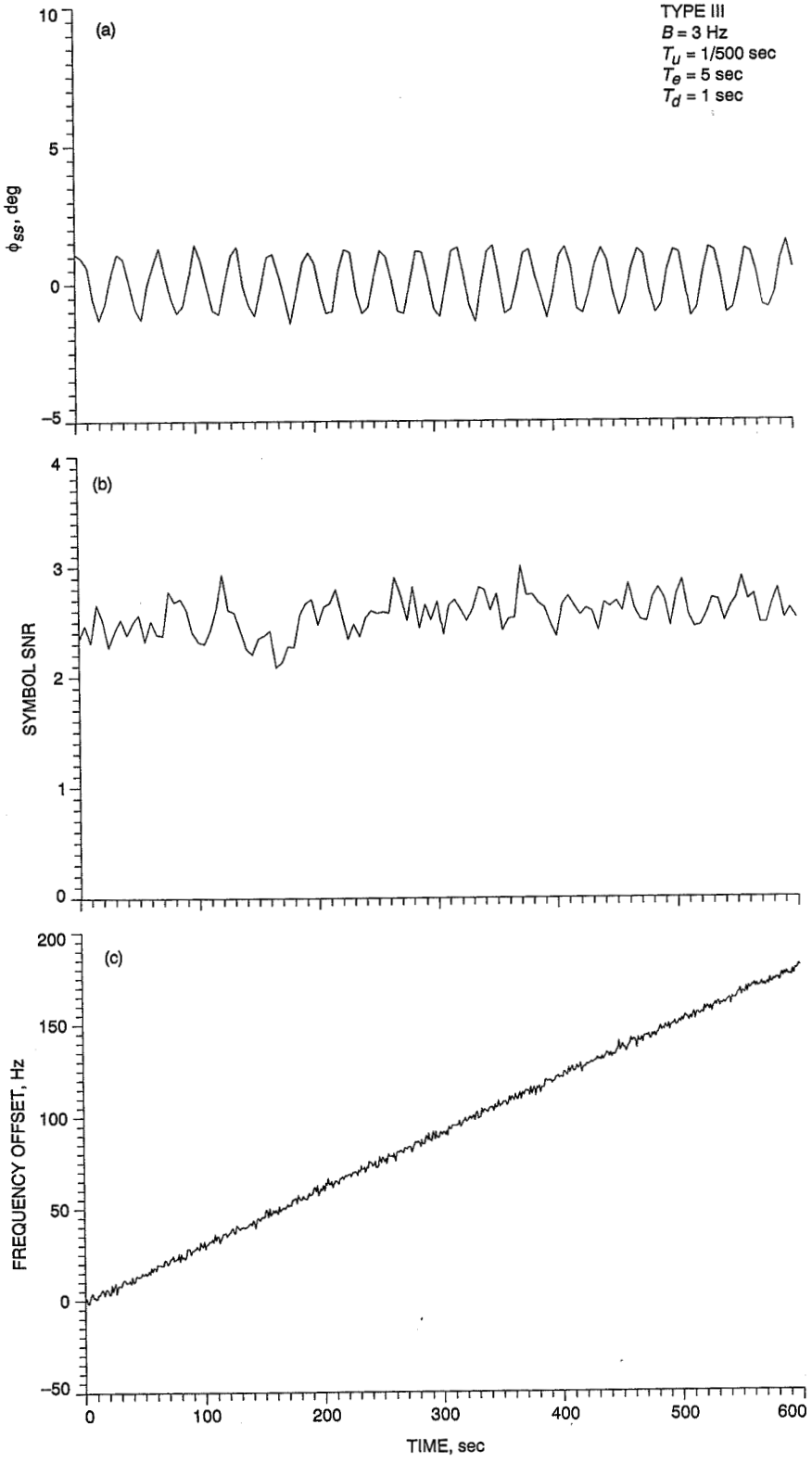


Fig. 7. Voyager 2 test results on DOY 293: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

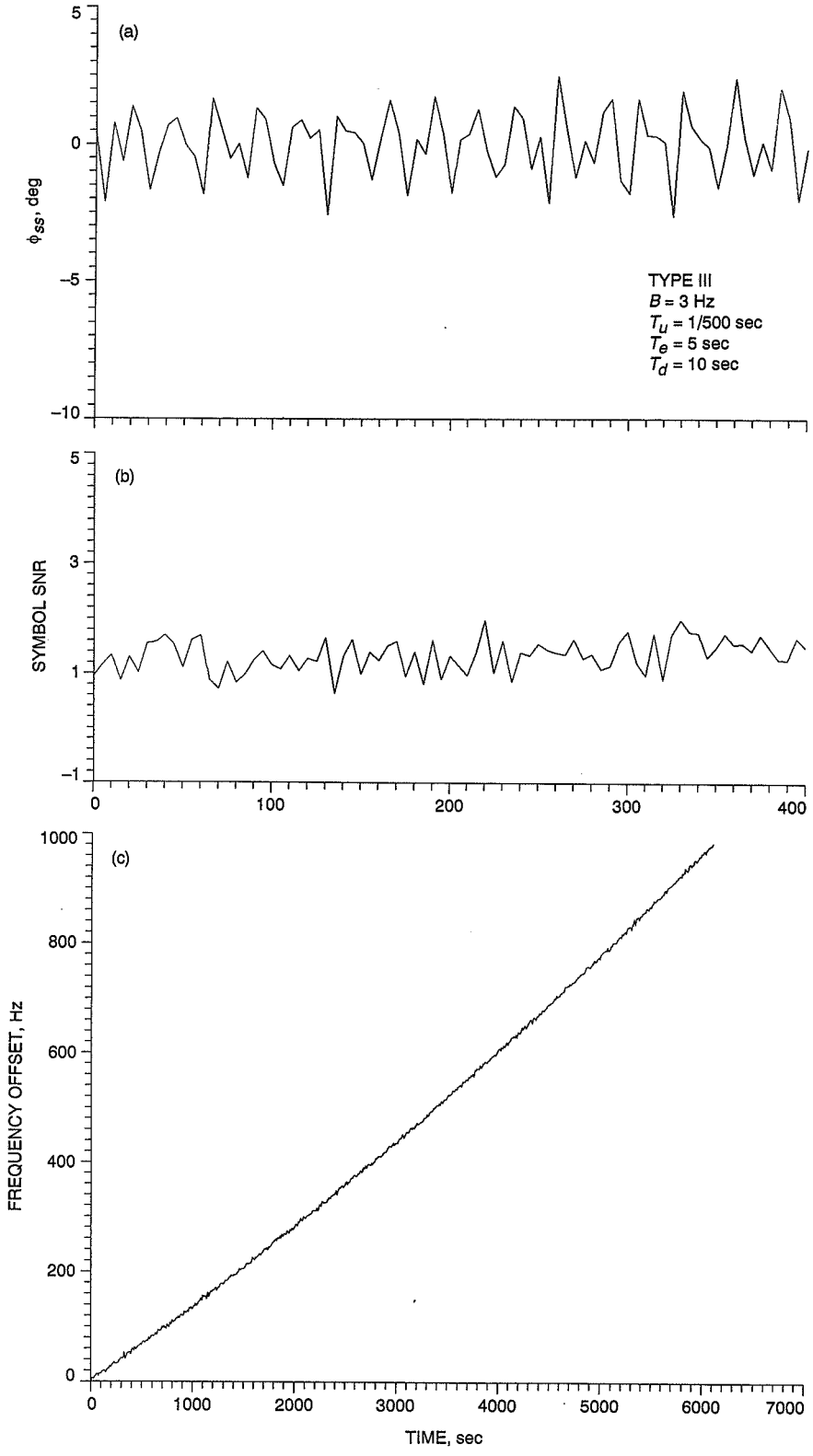


Fig. 8. ICE test results on DOY 293: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

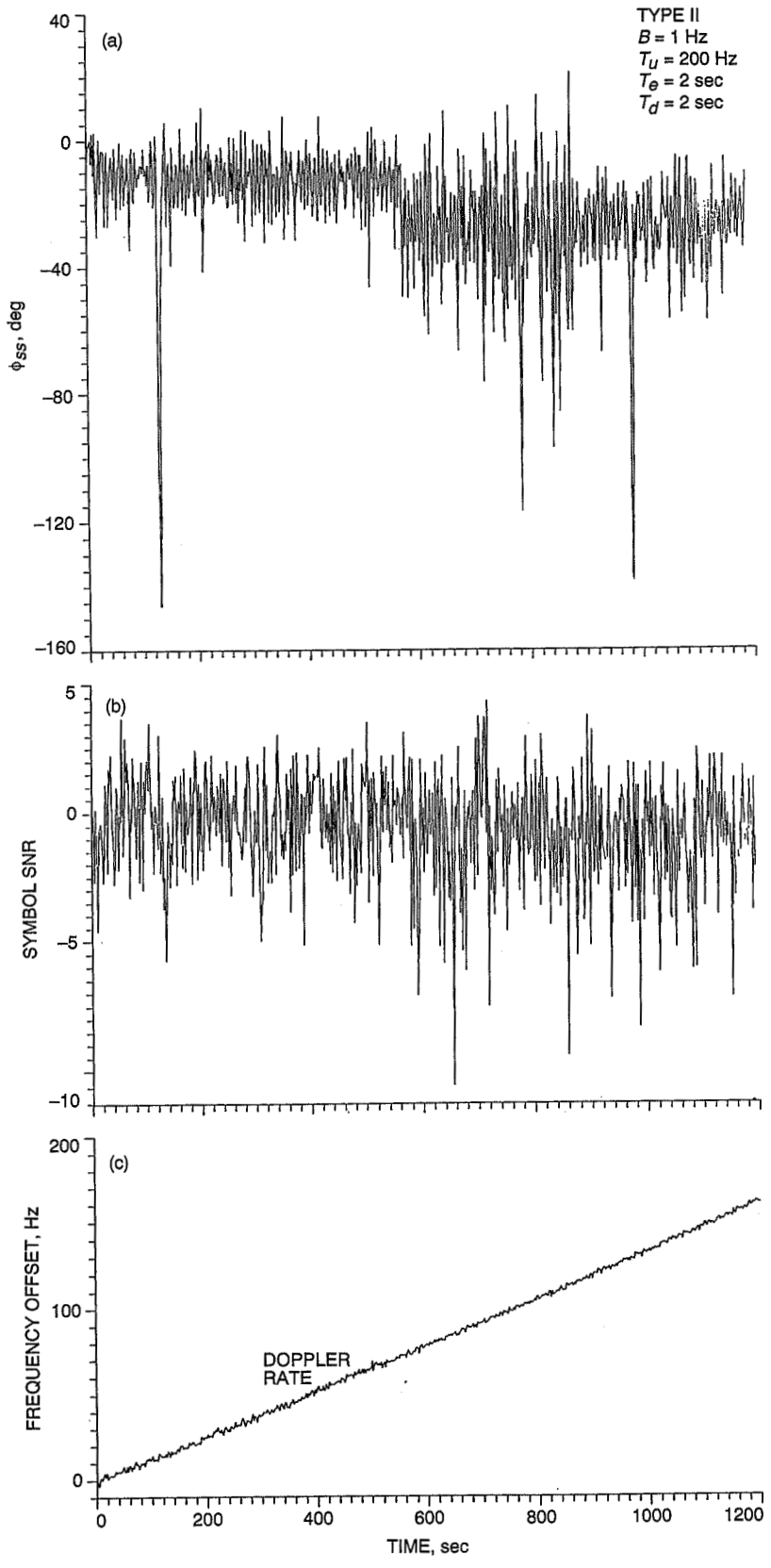


Fig. 9. Pioneer 10 test results on DOY 299: (a) ϕ_{ss}; (b) SSNR; and (c) ΔF.

Symbol Lock Detection in the ARX II and Block V Receivers

M. Shihabi and S. Hinedi
Communications Systems Research Section

B. Shah
Radio Frequency and Microwave Subsystems Section

The performances of five symbol-lock detectors are compared in this article. These detectors are the square-law detector with overlapping (SQOD) and nonoverlapping integrators, the absolute-value detectors with overlapping and nonoverlapping integrators, and the signal-power estimator detector (SPED). The analysis considers various scenarios in which the observation interval is much larger than or equal to the symbol-synchronizer loop bandwidth, and which have not been considered in previous analyses. Also, the case of threshold setting in the absence of signal is considered.

It is shown that the SQOD outperforms all others when the threshold is set in the presence of a signal, independent of the relationship between loop bandwidth and observation period. On the other hand, the SPED outperforms all others when the threshold is set in the presence of noise only.

I. Introduction

The Advanced Receiver II (ARX II) and Block V Receiver currently under development [1] use phase-locked loops (PLL's) to track the carrier, subcarrier, and symbol phase. Like most coherent receivers, the ARX II and Block V rely on lock detectors to provide the lock status of its PLL's. Since carrier, subcarrier, and symbol synchronization need to be achieved before any "meaningful" symbol detection can be initiated, lock detectors play a vital role in the final decision of accepting or rejecting the detected symbols. During operation, a loop is assumed to be locked when its lock indicator consistently has a positive status. The carrier and subcarrier lock detectors used in the ARX II and Block V Receiver have already been analyzed [2,3]. This article analyzes five candidate symbol-lock detectors for these receivers.

The detectors considered in this article are divided into two groups. The detectors in the first group process the outputs of two overlapping integrators, whereas those in the second group use two contiguous outputs of a single integrator. The first class contains two well-known lock detectors that use either squaring or absolute-value operations on the integrator outputs. The second class consists of three lock detectors, two of which use the same mathematical operations as those in the first class, while

the third detector functions as a signal power estimator. The five schemes are compared based on the lock-detection probability as a function of the symbol SNR for a given false-alarm probability and a fixed observation interval.

Although symbol-lock detection has been addressed before [4,5], the analyses have neglected the interdependence between symbol synchronizer bandwidth and lock detector bandwidth. The symbol synchronizer bandwidth refers to the one-sided loop noise bandwidth B_L of the digital data-transition tracking loop [6] used in the ARX II and Block V. The lock detector bandwidth is defined as the frequency at which the lock detector provides a status, being in- or out-of-lock. For example, the lock detectors considered here indicate loop status once every M symbols. Consequently, the bandwidth of these detectors is $1/(MT)$, where T is the symbol duration. The probability of false alarm P_{fa} is defined in two ways. In the classical sense, it is defined as the probability of declaring a signal (or target, as in radar applications) to be present when it is not present. In deep space applications, however, it is more appropriate to define P_{fa} as the probability of declaring a loop to be in-lock when it is out-of-lock. That is, P_{fa} is the probability of declaring the timing error to be "zero" (in-lock) when the loop is slipping cycles and operating with a nonzero timing error (out-of-lock).

In Section IV, the false alarm rate is shown to be drastically different, depending on the definition used. In addition, when the loop is slipping cycles, the false alarm rate is shown to depend strongly on the ratio of the lock detector bandwidth to the symbol loop bandwidth. For example, when the loop is slipping and $1/B_L \ll MT$, the lock detectors operate with acceptable false alarm rates because there are several uncorrelated samples of the timing error τ within the MT-sec decision interval. On the other hand, when $1/B_L \gg MT$, the false alarm rates are unacceptable because the timing error is constant over several decision intervals (see Section IV). Note that a good rule of thumb is to assume that the loop provides uncorrelated phase estimates every $1/B_L$ sec. As a result, the symbol timing error at time t_i is uncorrelated with the symbol timing error at time t_j, when $|t_i - t_j| \geq 1/B_L$. This article considers the special cases of $1/B_L = MT$ and $1/B_L = T$. The first case is analyzed and simulated, whereas the second case is simulated, but not analyzed. When the threshold is adjusted in the presence of noise only, the performance can be derived from the previous analysis by setting the signal amplitude to zero.

The description of the various lock detectors is presented in Section II. The general analysis and summary of the theoretical results are presented in Section III. The discussion of the theoretical, as well as the simulated, results is carried out in Section IV, followed by the conclusion in Section V. In Appendices A through F, some of the mathematical details are provided.

II. Description of the Lock Detectors

Figure 1 is a block diagram showing the signal processing functions common to the lock detectors surveyed in this article. The received signal is assumed to have been mixed with perfect carrier and subcarrier references so that the input to the lock detectors is a baseband signal of the form

$$r(t) = Ad(t) + n(t) \tag{1}$$

where A is the signal amplitude and A^2 is the received data power with

$$d(t) = \sum_{k=-\infty}^{\infty} d_k p(t - kT) \tag{2}$$

and where $n(t)$ is the additive white Gaussian noise process with a single-sided power spectral density (PSD) N_0 (watt/Hz). The data symbol d_k takes on the value ± 1 equally likely, and $p(t)$ is the received pulse shape of duration T sec. For comparison purposes, only the nonreturn-to-zero pulse is considered in the analysis, but the results can be extended to any pulse shape. The receiver is assumed to have perfect knowledge of T, but not the symbol epoch—i.e., the receiver has perfectly estimated the symbol rate, but not necessarily the start and end of the symbols.

The signal processing functions for the "lock detector" block in Fig. 1 depend on the detector being implemented. Its output X_k is at the symbol rate and, typically, many samples of X_k are averaged to obtain the decision statistic Y. If Y is greater than the threshold δ, the loop is declared to be in-lock, otherwise it is declared to be out-of-lock.

The parameter τ in Fig. 1 is the phase error between the symbol phase and the phase estimate provided by the symbol synchronizer. The in-lock case is analyzed by setting the timing error τ to zero. In practice, the error is not identically zero, but it is a very small value. When there is a signal present, the out-of-lock model for τ depends on the relation between B_L and $1/(MT)$. When $B_L = 1/(MT)$, τ is modelled as an unknown constant over a decision interval (MT sec), but one that is independent and uniformly

distributed from one decision interval to the next. Alternatively, when $B_L = 1/T$, the timing error is modelled as a constant over a symbol interval (T sec), but one that is independent and uniformly distributed from symbol to symbol. In this case, if the decision time $MT \gg T$ (as it usually is), each decision statistic encompasses the entire range of τ. When there is no signal present, the model of τ is irrelevant because the out-of-lock performance is independent of τ.

The detectors considered in this article are the square-law detector with overlapping (SQOD) and nonoverlapping (SQNOD) integrators, the absolute-value detectors with overlapping (AVOD) and nonoverlapping (AVNOD) integrators, and the signal-power estimator detector (SPED). The SQOD is depicted in Fig. 2(a). The input signal $r(t)$ is integrated over two symbol periods: one inphase with the estimated symbol interval and the other staggered by half a symbol duration. The resulting inphase and quadraphase samples I_k and Q_k are correlated due to the overlapping intervals. The AVOD detector replaces the squaring operations in Fig. 2(a) with absolute-value operations. Hence, its I and Q samples are also correlated. The SQNOD detector processes two contiguous outputs of a single integrator, as shown in Fig. 2(b). As before, replacing the squaring operations with absolute values yields the detector's counterpart, the AVNOD. The integrator outputs in this case are uncorrelated because the intervals are disjointed. The SPED detector is shown in Fig. 2(c). The SPED detector is considered because it is already part of the split-symbol moments' symbol-to-noise-ratio (SNR) estimator [7], the SNR estimator used in the ARX II and Block V. The inphase and quadraphase outputs of the SPED are obtained, respectively, by integrating the received signal over the first and second halves of a symbol. Since the noise in the first and second halves are independent, taking the product of I and Q and averaging over many symbols provides an estimate proportional to signal power.

III. Performance Analysis

The in-lock and out-of-lock performances of the lock detectors are derived in this section. The in-lock performance is measured in terms of the probability of declaring the loop to be locked when there is no timing error—that is, the probability that the decision statistic Y in Fig. 1 is greater than the threshold δ when $\tau = 0$. Note that $\tau = 0$ or no phase-tracking error is equivalent to setting the symbol synchronizer loop SNR to infinity. The degradation in detection probability due to timing jitter (a noninfinite loop SNR) is minimal and has been addressed in the case of carrier lock detectors [2]. The out-of-lock performance is measured by the probability of false alarm—the probability of declaring the loop to be locked when it is not locked. The out-of-lock performance in the presence of a signal is analyzed for the case $B_L = 1/(MT)$. The simulation results for $B_L = 1/T$ are presented in the next section.

Note that when $B_L = 1/T$, the timing error is independent from one symbol to the other, and the decision is made after averaging many, say M, symbols. On the other hand, when $B_L = 1/(MT)$, τ is an unknown constant during a decision interval, and is independent from one decision to the next. Setting $M = 1$ in the latter case would imply a decision for every symbol, which is fundamentally different from the case where $B_L = 1/T$ and the decision is made by using M symbols. Hence, the performance when $B_L = 1/T$ cannot be derived from the case of $B_L = 1/(MT)$ simply by setting $M = 1$. The out-of-lock performance when there is no signal present is also analyzed. Only the final equations are shown with the derivations detailed in the various appendices. In all cases, the decision statistic can be expressed as

$$Y = \frac{1}{M}\sum_{k=1}^{M} X_k \tag{3}$$

where the random variable X_k is peculiar to each detector. When the timing offset $\tau = 0$, the adjacent samples X_k and X_{k+1} are correlated in the first two detectors. Whereas for the remaining three detectors, they are uncorrelated. In all cases, the random variable X_k is not Gaussian due to the nonlinear operations on I_k and Q_k. For large values of M, the random variable Y is modelled as Gaussian due to the central limit theorem (CLT). The theorem applies to the sum of correlated random variables when none of the variables being summed dominates over the others [8]. This model for Y is justified by simulation results.

The probability of lock detection is the probability that the Gaussian random variable Y surpasses the threshold δ. Hence, it is given by

$$P_D = \int_{\delta}^{\infty} \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] dy \tag{4}$$

where μ_Y and σ_Y^2 are the mean and variance of Y when τ is exactly zero. By using the definition of the error function

$$\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \tag{5}$$

one has

$$P_D = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{\delta - \mu_Y}{\sqrt{2}\sigma_Y}\right) \tag{6}$$

or

$$P_D = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{\delta}{\sqrt{2}\sigma_Y}\sqrt{\frac{SNR_D}{2}}\right) \tag{7}$$

where SNR_D denotes the detector SNR defined by

$$SNR_D \triangleq \left(\frac{\mu_Y}{\sigma_Y}\right)^2 \tag{8}$$

The threshold δ is chosen to maintain a fixed probability of false alarm. The probability of false alarm is the probability that the out-of-lock decision statistics do not surpass the threshold. Hence, it is given by

$$P_{fa} = \int_{-\infty}^{\delta} f_o(y)dy \tag{9}$$

where $f_o(y)$ is the out-of-lock density of Y. The threshold δ is computed by solving Eq. (9) for a fixed P_{fa}. When there is a signal present and $B_L = 1/(MT)$, the statistic Y is no longer Gaussian and $f_o(y)$ must be obtained numerically or by simulation, as shown in Section IV. When there is no signal present, the CLT can be invoked and the out-of-lock decision statistic can be modelled as Gaussian. This model is verified by simulations in Section IV. In this case, Eq. (9) can be written as

$$P_{fa} = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{\delta - \mu_{Y_o}}{\sqrt{2}\sigma_{Y_o}}\right) \tag{10}$$

where μ_{Y_o} and σ_{Y_o} are the out-of-lock mean and variance of the decision statistic Y. The threshold δ is given by

$$\delta = \sqrt{2}\sigma_{Y_o}\gamma + \mu_{Y_o} \tag{11}$$

where $\gamma = \mathrm{erf}^{-1}(1 - 2P_{fa})$. Substituting Eq. (11) into Eq. (6) relates the probability of detection to the no-signal (classical) false alarm rate, namely,

$$P_D = \frac{1}{2} - \frac{1}{2}\,\mathrm{erf}\left(\frac{\sqrt{2}\sigma_{Y_o}\gamma + \mu_{Y_o} - \mu_Y}{\sqrt{2}\sigma_Y}\right) \tag{12}$$

The next five subsections derive the in-lock and out-of-lock mean and variance for all five schemes.

A. Square-Law Lock Detector With Overlapping Intervals (SQOD)

The SQOD detector is shown in Fig. 2(a). For the input given by Eq. (1), the inphase integrator output is given by

$$I_k = \int_{kT+\tau}^{(k+1)T+\tau} r(t)dt$$

$$= d_k A(T-\tau) + d_{k+1}A\tau + N_1(k) + N_2(k) \tag{13}$$

and the quadraphase integrator output is given by

$$Q_k = \int_{(k+\frac{1}{2})T+\tau}^{(k+\frac{3}{2})T+\tau} r(t)dt$$

$$= \begin{cases} d_k A\left(\frac{T}{2}-\tau\right) + d_{k+1}A\left(\frac{T}{2}+\tau\right) \\ \quad +N_2(k) + N_1(k+1) & 0 \le \tau < \frac{T}{2} \\ d_{k+1}A\left(\frac{3T}{2}-\tau\right) + d_{k+2}A\left(\tau - \frac{T}{2}\right) \\ \quad +N_2(k) + N_1(k+1) & \frac{T}{2} \le \tau < T \end{cases} \tag{14}$$

where τ is limited to the interval $[0,\ T]$, and where

$$N_1(k) = \int_{(kT+\tau)}^{(k+\frac{1}{2})T+\tau} n(t)dt \tag{15}$$

and

$$N_2(k) = \int_{(k+\frac{1}{2})T+\tau}^{(k+1)T+\tau} n(t)dt \tag{16}$$

Since $n(t)$ is a white Gaussian process with one-sided PSD N_o, the N_i's are independent Gaussian random variables with a mean of zero and a variance of $\sigma_n^2 = (N_oT)/4$. Summing M of the samples $X_k = I_k^2 - Q_k^2$ yields Y. From Appendix A, the in-lock mean and variance of Y are given by

$$\mu_Y = \frac{\eta_s NoT}{2} \tag{17}$$

and

$$\sigma_Y^2 = \left(\frac{N_0T}{M}\right)^2 \left[M\left(\frac{\eta_s^2}{4} + 2\eta_s + \frac{3}{4}\right) + 2(M-1)\left(-\frac{\eta_s}{2} - \frac{1}{8}\right)\right] \quad (18)$$

The out-of-lock mean and variance, when there is a signal present and τ is an unknown constant over a decision interval $B_L = 1/(MT)$, are given by

$$\mu_{Y_o} = 0.0 \quad (19)$$

$$\sigma_{Y_o}^2 = \left(\frac{N_0T}{M}\right)^2 \left[M\left(\frac{31\eta_s^2}{120} + \frac{41\eta_s}{24} + \frac{3}{4}\right) + 2(M-1) \times \left(\frac{23\eta_s^2}{320} - \frac{23\eta_s}{48} - \frac{1}{8}\right) + (M-1)(M-2)\frac{\eta_s^2}{12}\right] \quad (20)$$

where η_s denotes the symbol signal-to-noise ratio and is defined as

$$\eta_s \triangleq \frac{A^2T}{N_0} \quad (21)$$

Setting $A = 0$ in Eq. (21) and substituting the result in Eqs. (19) and (20) yields the out-of-lock mean and variance in the no-signal case. Hence, the no-signal mean is zero, but the no-signal variance is given by

$$\sigma_{Y_o}^2(\eta_s = 0) = \left(\frac{N_0T}{M}\right)^2 \left[M\frac{3}{4} - 2(M-1)\frac{1}{8}\right] \quad (22)$$

B. Absolute-Value Lock Detector With Overlapping Intervals (AVOD)

For the AVOD detector, Fig. 2(a) with absolute values instead of squares, the expressions for I_k and Q_k given by Eqs. (13) and (14) are still valid, but now $X_k = |I_k| - |Q_k|$. From Appendix B, the in-lock mean and variance are given by

$$\mu_Y = \left(\sqrt{\frac{N_0T}{4}}\right)\left[\frac{\exp(-\eta_s) - 1}{\sqrt{\pi}} + \sqrt{\eta_s}\,\mathrm{erf}(\sqrt{\eta_s})\right] \quad (23)$$

and

$$\sigma_Y^2 = \frac{1}{M^2}[M\ \mathrm{Var}(X_k) + M(M-1)\ \mathrm{Cov}(X_k, X_{k+1})] \quad (24)$$

where

$$\mathrm{Var}(X_k) = (N_0T)\left\{\frac{3\eta_s}{2} + 1 - \eta_s[F_1(\eta_s) + F_2(\eta_s)] - \frac{1}{4\pi}[\exp(-2\eta_s) - 2\exp(-\eta_s) + 1] - \frac{\eta_s}{4}\mathrm{erf}^2(\sqrt{\eta_s}) - \frac{1}{2}\sqrt{\frac{\eta_s}{\pi}}\mathrm{erf}(\sqrt{\eta_s})[\exp(-\eta_s) - 1]\right\} \quad (25)$$

and

$$\mathrm{Cov}(X_k, X_{k+1}) = (N_0T)\left\{\frac{\exp(-2\eta_s) + \exp(-\eta_s)}{2\pi} + \frac{\eta_s}{2}\mathrm{erf}^2(\sqrt{\eta_s}) + \left[\exp(-\eta_s) + \frac{1}{2}\right]\mathrm{erf}(\sqrt{\eta_s}) \times\sqrt{\frac{\eta_s}{\pi}} - \frac{\eta_s}{2}[F_1(\eta_s) + F_2(\eta_s)]\right\} \quad (26)$$

The out-of-lock mean and variance, when there is a signal and τ is constant over M symbols, are given by

$$\mu_{Y_o} = 0.0 \quad (27)$$

and

$$\sigma_{Y_o}^2 = \frac{1}{M^2}[M\ \mathrm{Var}_o(X_k) + 2(M-1)\ \mathrm{Cov}_o(X_k, X_{k+1}) + (M-1)(M-2)\ \mathrm{Cov}_o(X_k, X_{k+2})] \quad (28)$$

where

$$\mathrm{Var}_o(X_k) = (N_0T)\left\{\frac{4\eta_s}{3} + 1 - \eta_s\left[\frac{3}{4}F_1(\eta_s) + G_1(\eta_s) + \frac{H_1(\eta_s) + H_2(\eta_s) + H_3(\eta_s)}{2}\right]\right\} \quad (29)$$

$$\mathrm{Cov}_o(X_k, X_{k+1}) = (N_0T)\left\{\frac{\exp(-2\eta_s)}{2\pi} + \left(\frac{\eta_s}{2} + \frac{1}{8}\right)\right.$$

$$\times \mathrm{erf}^2(\sqrt{\eta_s}) + \exp(-\eta_s)\mathrm{erf}(\sqrt{\eta_s})\left(\sqrt{\frac{\eta_s}{\pi}} + \frac{1}{8\sqrt{\pi\eta_s}}\right)$$

$$+\frac{\eta_s}{4}[F_3(\eta_s) + G_3(\eta_s) + H_4(\eta_s) - G_4(\eta_s) - H_5(\eta_s)$$

$$\left.-\frac{3}{2}F_1(\eta_s) - G_1(\eta_s) - 2G_2(\eta_s) - 2H_2(\eta_s)]\right\} \qquad (30)$$

$$\mathrm{Cov}_o(X_k, X_{k+j}) = \left(\frac{N_0T\eta_s}{4}\right)[F_3(\eta_s) + G_3(\eta_s) + H_4(\eta_s)$$

$$- 2G_4(\eta_s) - 2H_5(\eta_s)] \qquad \text{for j} \geq 2 \qquad (31)$$

The functions F_i, G_i, and H_i in Eqs. (28) through (31) are defined in Appendix B and plotted versus η_s in Fig. B-1. Setting $\eta_s = 0$ in Eqs. (28) through (31) yields the out-of-lock statistics in the no-signal case. The no-signal mean is zero, but the variance is given by

$$\sigma^2_{Y_o}(\eta_s = 0) = \left(\frac{N_0T}{M^2}\right)\left[M + 2(M-1)\left(\frac{1}{2\pi}\right)\right] \qquad (32)$$

C. Square-Law Lock Detector With Nonoverlapping Intervals (SQNOD)

The SQNOD detector is shown in Fig. 2(b). For the input of Eq. (1), The inphase and quadraphase integrator outputs are given by

$$I_k = \int_{(k+\frac{1}{4})T+\tau}^{(k+\frac{3}{4})T+\tau} r(t)dt$$

$$= \begin{cases} d_k\frac{AT}{2} + N_1(k) & 0 \leq \tau < \frac{T}{4} \\ d_kA\left(\frac{3T}{4} - \tau\right) & \\ \quad +d_{k+1}A\left(\tau - \frac{T}{4}\right) + N_1(k) & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \qquad (33)$$

and

$$Q_k = \int_{(k-\frac{1}{4})T+\tau}^{(k+\frac{1}{4})T+\tau} r(t)dt$$

$$= \begin{cases} d_{k-1}A\left(\frac{T}{4} - \tau\right) & \\ \quad +d_kA\left(\frac{T}{4} + \tau\right) + N_2(k) & 0 \leq \tau < \frac{T}{4} \\ d_k\frac{AT}{2} + N_2(k) & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \qquad (34)$$

The noises $N_1(k)$ and $N_2(k)$ are given by Eqs. (15) and (16) after changing the integration limits to those in Eqs. (33) and (34). As a result, they are independent Gaussian random variables with zero mean and variance $\sigma_n^2 = (N_0T)/4$. The sample X_k is the difference of the squares (i.e., $X_k = I_k^2 - Q_k^2$). From Appendix D, the in-lock mean and variance of Y are given by

$$\mu_Y = \frac{\eta_s N_0 T}{8} \qquad (35)$$

and

$$\sigma_Y^2 = \left(\frac{N_0^2T^2}{M}\right)\left(\frac{\eta_s^2 + 24\eta_s + 16}{64}\right) \qquad (36)$$

For the case of false lock with the signal present and where τ is an unknown constant over M symbols, one obtains

$$\mu_{Y_o} = 0.0 \qquad (37)$$

and

$$\sigma^2_{Y_o} = \left(\frac{N_0^2T^2}{M^2}\right)$$

$$\times\left[M\left(\frac{\eta_s^2 + 25\eta_s + 15}{60}\right) + M(M-1)\frac{\eta_s^2}{120}\right] \qquad (38)$$

When there is no signal present, the out-of-lock mean is zero and the variance is given by setting $\eta_s = 0$ in Eq. (38). Consequently, the no-signal out-of-lock variance is

$$\sigma^2_{Y_o}(\eta_s = 0) = \left(\frac{N_0^2T^2}{M}\right)\left(\frac{15}{60}\right) \qquad (39)$$

D. Absolute-Value Lock Detector With Disjoint Intervals (AVNOD)

This detector is the same as the SQNOD with the squaring operations replaced by absolute value operations. Hence, Eqs. (33) and (34) for I_k and Q_k are valid, but now

$X_k = |I_k| - |Q_k|$. From Appendix E, the in-lock statistics for Y are given by

$$\mu_Y = \sqrt{\frac{N_0T}{4}}\left[\frac{\exp\left(-\frac{\eta_s}{2}\right)-1}{\sqrt{2\pi}} + \frac{\sqrt{\eta_s}}{2}\text{erf}\left(\sqrt{\frac{\eta_s}{2}}\right)\right] \quad (40)$$

and

$$\sigma_Y^2 = \left(\frac{N_0T}{M}\right)\left\{\frac{3\eta_s}{8} + \frac{1}{2} - \frac{1}{8\pi}\left[1 + 5\exp(-\eta_s)\right.\right.$$

$$\left. + 2\exp\left(-\frac{\eta_s}{2}\right)\right] - \sqrt{\frac{\eta_s}{32\pi}}\text{erf}\left(\sqrt{\frac{\eta_s}{2}}\right)$$

$$\left.\times\left[5\exp\left(-\frac{\eta_s}{2}\right)+1\right] - \frac{5\eta_s}{16}\text{erf}^2\left(\sqrt{\frac{\eta_s}{2}}\right)\right\} \quad (41)$$

The out-of-lock mean and variance when there is a signal present and τ is an unknown constant over MT sec are given by

$$\mu_{Y_o} = 0.0 \quad (42)$$

and

$$\sigma_{Y_o}^2 = \frac{1}{M^2}[M\ \text{Var}_o(X_k) + M(M-1)\ \text{Cov}_o(X_k, X_{k+1})] \quad (43)$$

$$\text{Var}_o(X_k) = (N_0T)\left[\frac{5\eta_s}{12} + \frac{1}{2} - \frac{3}{4\pi}\exp(-\eta_s)\right.$$

$$-\text{erf}^2\left(\sqrt{\frac{\eta_s}{2}}\right)\left(\frac{1+3\eta_s}{8}\right) - \exp\left(-\frac{\eta_s}{2}\right)$$

$$\left.\times\ \text{erf}\left(\sqrt{\frac{\eta_s}{2}}\right)\left(\frac{3}{2}\sqrt{\frac{\eta_s}{2\pi}} + \frac{1}{\sqrt{32\pi\eta_s}}\right)\right] \quad (44)$$

$$\text{Cov}_o(X_k, X_{k+1}) = (N_0T)\left\{-\frac{1}{16}\text{erf}^2\left(\sqrt{\frac{\eta_s}{2}}\right) - \exp\left(-\frac{\eta_s}{2}\right)\right.$$

$$\left.\times\ \text{erf}\left(\sqrt{\frac{\eta_s}{2}}\right)\left(\frac{1}{8\sqrt{2\pi\eta_s}}\right) + \frac{\eta_s}{2}Z(\eta_s)\right\} \quad (45)$$

where the function Z is defined in Appendix E and plotted in Fig. B-2. For the out-of-lock case with no signal, the mean is zero and the variance is obtained by setting $\eta_s = 0$ in Eqs. (43) through (45). Hence, the out-of-lock variance is given by

$$\sigma_{Y_o}^2(\eta_s = 0) = \frac{N_0T}{M}\left(\frac{1}{2} - \frac{3}{4\pi}\right) \quad (46)$$

E. Signal-Power Estimator Lock Detector (SPED)

This detector is shown in Fig. 2(c). Denote the integrations over the first half of the assumed symbol interval as I_k and the second half as Q_k. Then, the I_k and Q_k samples are given by

$$I_k = d_k\frac{AT}{2} + N_1(k) \quad (47)$$

and

$$Q_k = d_kA\left(\frac{T}{2} - \tau\right) + d_{k+1}A\tau + N_2(k) \quad (48)$$

and $X_k = I_kQ_k$. From Appendix F, the in-lock mean and variance of Y are

$$\mu_Y = \frac{\eta_sN_0T}{4} \quad (49)$$

and

$$\sigma_Y^2 = \left(\frac{N_0^2T^2}{M}\right)\left(\frac{2\eta_s+1}{16}\right) \quad (50)$$

The out-of-lock case with signal present has a mean and variance, when τ is constant over M symbols, given by

$$\mu_{Y_o} = \frac{\eta_sN_0T}{8} \quad (51)$$

and

$$\sigma_{Y_o}^2 = \frac{N_0^2T^2}{M^2}\left[M\left(\frac{5\eta_s^2+20\eta_s+12}{192}\right) + M(M-1)\frac{\eta_s^2}{192}\right] \quad (52)$$

As before, the out-of-lock variance in the no-signal case is given by

$$\sigma^2_{Y_o}(\eta_s = 0) = \left(\frac{N_0^2 T^2}{M}\right)\left(\frac{12}{192}\right) \tag{53}$$

IV. Discussion and Simulation Results

Digital simulation was used to verify the foregoing analysis. The first part of this section presents the results for the long-time constant case or $B_L = 1/(MT)$. Simulation results for the short-time constant case or $B_L = 1/T$ are presented in Subsection IV.B. The last subsection discusses the no-signal case.

A. Long-Time Constant, $B_L = 1/(MT)$

In the out-of-lock state, the symbol timing error τ is modelled as a constant over a decision interval (MT sec), but it is independent and uniformly distributed over the collection of all decision intervals. The timing error in the in-lock state is modelled as being zero. Although the special case where τ is constant over M symbols was analyzed for performance comparison purposes, it is not advisable to operate a practical system under these conditions due to unacceptable false alarm rates. This case has higher than usual false alarm rates because the decision statistic for small values of τ is not significantly different from the statistic for $\tau = 0$. As a result, the out-of-lock states corresponding to small values of τ are frequently declared to be in-lock because they are mistaken for the case when $\tau = 0$. This problem can be ameliorated by lengthening the observation time relative to the time constant of τ (i.e., shortening the time constant of τ). In practice, it is recommended that the observation time be at least ten times longer than the time constant of τ.

As noted in Section II, the out-of-lock density function for Y in this case is not Gaussian. Consider the decision statistic Y when the loop is out-of-lock and τ is constant over M symbols. In general, it can be written as

$$X_k = s_k(\tau) + n_k + s_k(\tau)n_k \tag{54}$$

where (in all five cases) the signal term s_k is random and uniformly distributed because τ is a uniformly distributed random variable. The density of the noise n_k depends on the detector being implemented. Summing M samples of X_k (where X_k is at the symbol rate in all cases) yields the decision variable Y. Since τ is constant over the sum, at a high SNR (i.e., for strong signal levels) the density function of Y approaches a uniform distribution as shown in Fig. 3(a). However, at a low SNR the noise term dominates and the density of Y is Gaussian due to the central limit theorem, as shown in Fig. 3(b). The density in Fig. 3 was obtained via numerical integration, as well as simulation. Both methods are seen to agree very well. The numerical method computed the density function of Y and $f(y)$, by averaging over τ the conditional probability density function $f(y/\tau)$. The latter is Gaussian with mean and variance where both are functions of τ. The simulation method computed the histogram of Y and then set $f(y) = [P(y - \Delta \leq Y \leq y + \Delta)]/\Delta$, where Δ is the size of a histogram bin. The histograms were generated by using 1,000,000 symbols, which corresponds to 10,000 decisions (Y's), since there are 100 symbols/decision.

Figure 4 compares the probability of detection performance for all five detectors for $M = 100$ and $P_{fa} = 0.25$. Note that the overlapping detectors SQOD and AVOD, which are identical except for the squaring and absolute value operations, have nearly identical performances. As expected, the AVOD is slightly better at a high SNR, whereas the SQOD is slightly better at a low SNR. The nonoverlapping detectors SQNOD and AVNOD also have nearly equal performance. Once again, the absolute value operation yields better results at higher SNR's. The SPED is better than the nonoverlapping detectors, but worse than the overlapping detectors. The probability of detection results in Fig. 4 change when P_{fa} or M change. For example, increasing the observation interval increases the detection probability because it increases the detector SNR (μ_Y^2/σ_Y^2). Accepting a higher false alarm rate increases the probability of detection because it lowers the threshold δ. In generating these curves, 50,000 symbols were simulated for each value of SNR. Since there are 100 symbols/decision, the detection probability for a given SNR is based on 500 decisions.

B. Short-Time Constant, $B_L = 1/T$

Here, for the out-of-lock state, the symbol timing error τ is modelled as a uniformly distributed random variable that changes independently from symbol to symbol. For this case, the probabilities of detection for all five detectors are computed by simulation for $M = 100$ and $P_{fa} = 10^{-2}$, and the threshold δ is set according to Eq. (11). The false alarm rate was verified by simulation. The results are plotted versus the symbol energy-to-noise ratio η_s in Fig. 5. In these computer simulations, the detection probability for a given SNR is based on 40,000 decisions.

The results show that the AVOD performs slightly better than SQOD at a high SNR, whereas the AVOD and SQOD seem to perform identically at a low SNR. The nonoverlapping detectors SQNOD and AVNOD also have nearly equal performance at a low SNR, but AVNOD performs about 1 dB better for values of the symbol SNR

higher than −4 dB. The SPED performs about 2 dB worse than the overlapping detectors and 3 dB better than the other two nonoverlapping detectors.

Also by simulation, the false-alarm rate that was used in setting the threshold was verified.

C. No-Signal Case

This scenario distinguishes between the case when there is no signal and when there is a signal and $\tau = 0$. Clearly, the out-of-lock statistic is Gaussian with a zero mean, and the in-lock statistic is Gaussian with a nonzero mean. Probability of detection results are compared in Fig. 6. Interestingly, the performances of the overlapping and nonoverlapping schemes are grouped together, but the SPED now has the best performance. The interdependence among P_D, P_{fa}, and M is the same as in the other two cases.

V. Conclusion

The performances of five symbol lock detectors are compared in this article. These detectors are the square-law detector with overlapping and nonoverlapping integrators, the absolute value detectors with overlapping and nonoverlapping integrators, and the signal-power estimator detector. The analysis considered various scenarios in which the observation interval is much larger than or equal to the symbol synchronizer loop bandwidth, and which have not been considered in previous analyses. Also, the case of threshold setting in the absence of signal was considered.

It is shown that the SQOD outperforms all others when the threshold is set in the presence of signal, independent of the relationship between loop bandwidth and observation period. On the other hand, the SPED outperforms all others when the threshold is set in the presence of noise only.

Acknowledgments

The authors would like to thank A. Mileant for proposing the idea of disjoint intervals. Also, the valuable comments made by Dr. W. J. Hurd are greatly appreciated.

References

[1] S. Hinedi, "A Functional Description of the Advanced Receiver II," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 131–149, February 15, 1990.

[2] A. Mileant and S. Hinedi, "Costas Loop Lock Detection in the Advanced Receiver," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 72–89, November 15, 1989.

[3] A. Mileant and S. Hinedi, "QPSK Loop Lock Detection in the Advanced Receiver," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 72–89, November 15, 1989.

[4] J. K. Holmes, *Coherent Spread Spectrum System*, New York: John Wiley and Sons, 1982.

[5] K. T. Woo, *Shuttle Bit Synch Lock Detector Performance*, TRW IOC No. SCTE-50-76-184/KTW, TRW Corporation, El Segundo, California, April 5, 1976.

[6] M. Simon, "An Analysis of the Steady-State Phase Noise Performance of a Digital Data-Transition Tracking Loop," *JPL Space Programs Summary 37-55*, vol. 3, Jet Propulsion Laboratory, Pasadena, California, pp. 54–62, February 28, 1969.

[7] M. Simon and A. Mileant, "SNR Estimation for the Baseband Assembly," *TDA Progress Report 42-85*, vol. January–March 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 118–125, May 15, 1986.

[8] D. Fraser, *Non-Parametric Methods in Statistics*, New York: John Wiley and Sons, 1957.

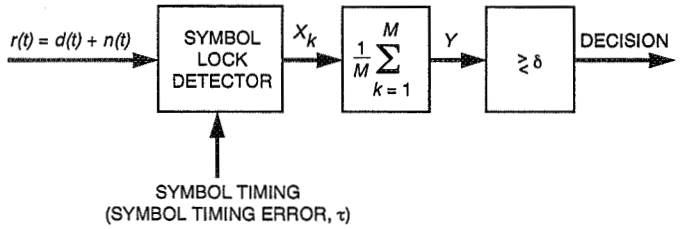


Fig. 1. Signal-processing functions common to all five lock detectors.

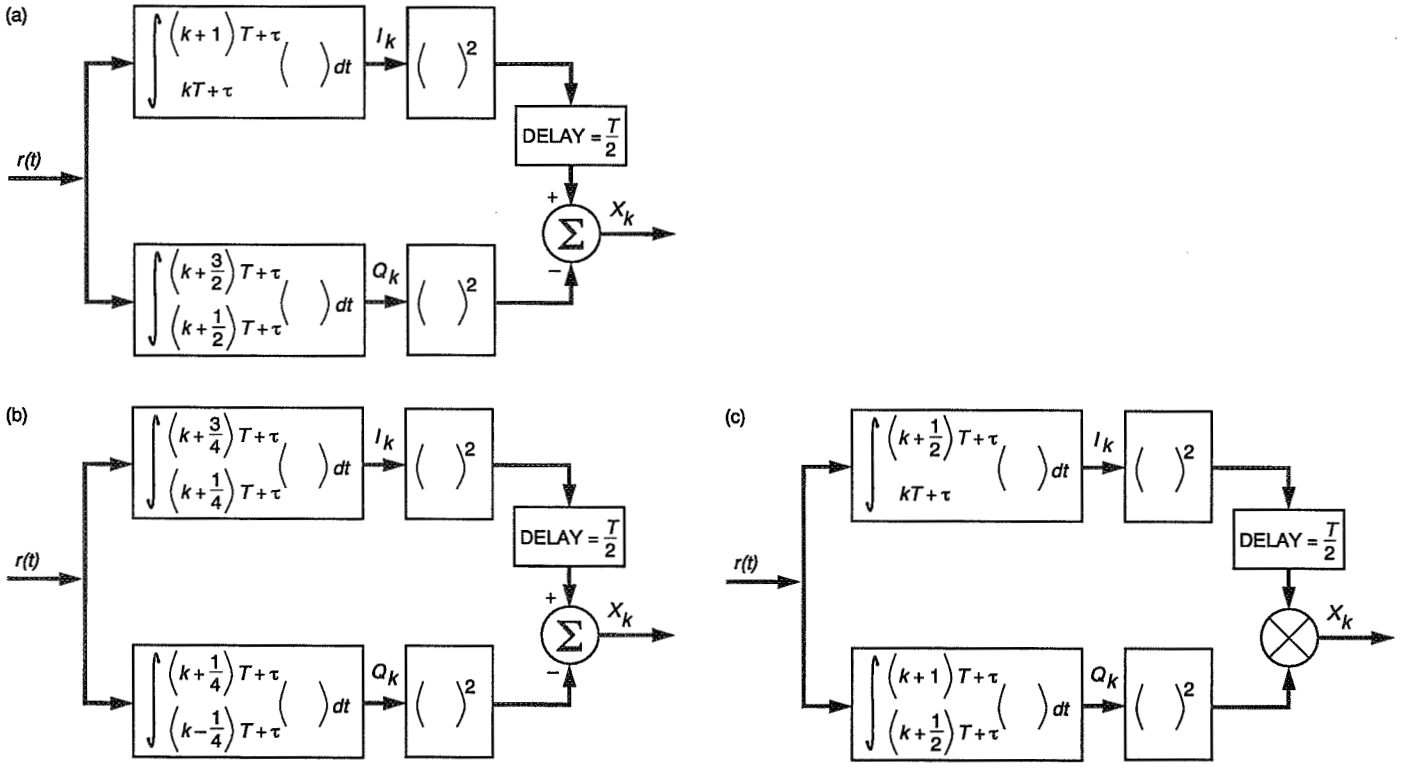


Fig. 2. Three of the detectors: (a) the SQOD, replacing $(\ )^2$ with 11 yields the absolute-value detector AVOD), (b) the SQNOD, replacing $(\ )^2$ with 11 yields the absolute-value detector AVNOD, and (c) the SPED.

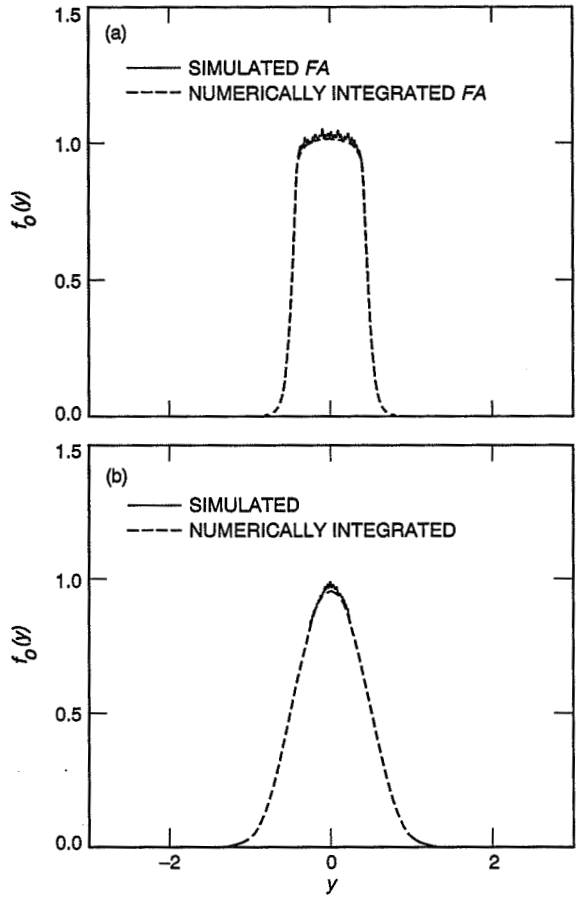


Fig. 3. The probability density function of Y for the SQOD, when the loop is out of lock, has a: (a) high SNR = 5 dB and (b) low SNR = −5 dB.

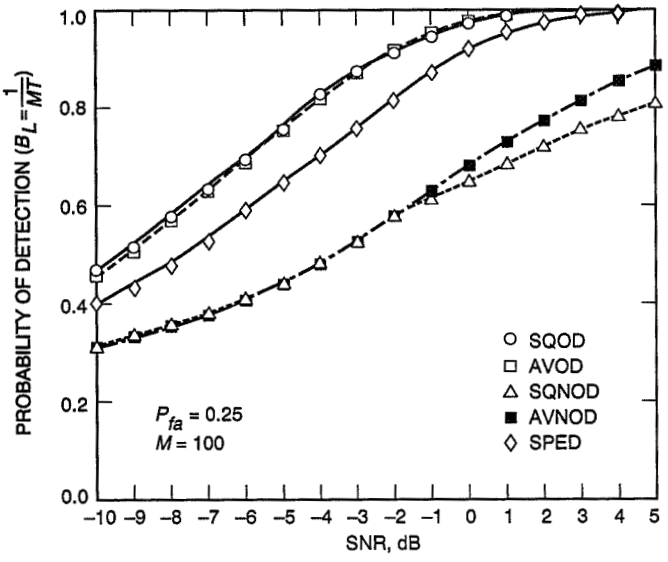


Fig. 4. The probability of detection versus SNR when τ is an unknown constant over a decision interval.

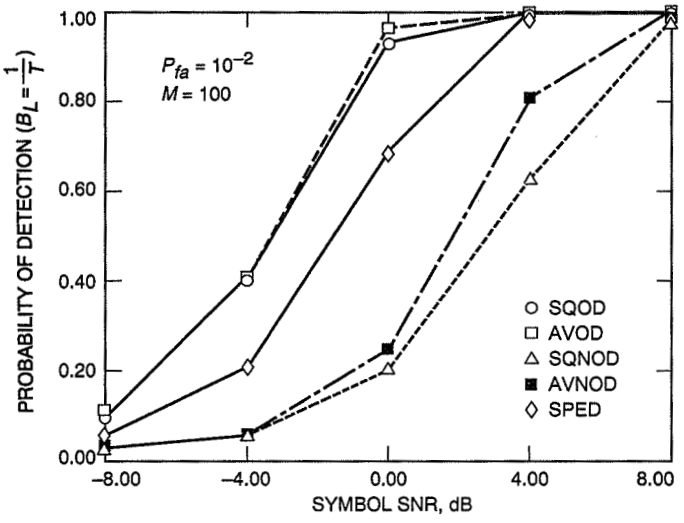


Fig. 5. The probability of detection versus SNR when τ is uniformly distributed and charging from symbol to symbol.

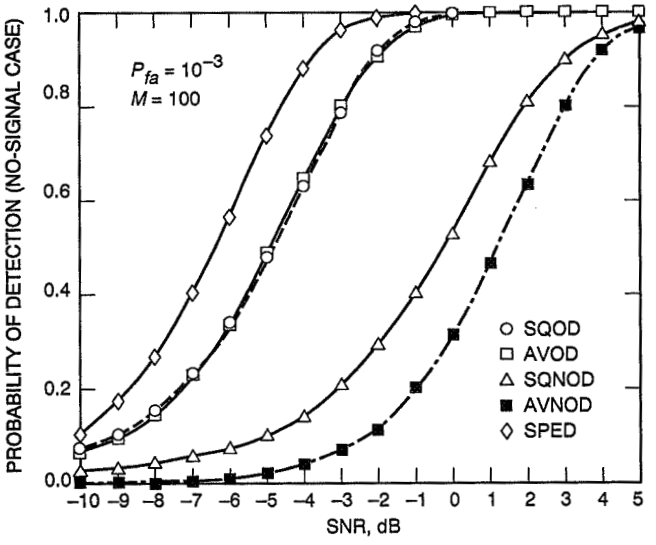


Fig. 6. The probability of detection versus SNR when the false alarm rate is computed in the absence of a signal.

Appendix A

Derivation of the Mean and Variance of the SQOD

The inphase and quadraphase integrator outputs are given by Eqs. (13) and (14), respectively. The output of the lock detector $X_k = I_k^2 - Q_k^2$. Consequently,

$$\mu_k = \mathcal{E}\{I_k^2\} - \mathcal{E}\{Q_k^2\} \tag{A-1}$$

where $\mu_k = \mathcal{E}\{X_k\}$.

$$\begin{aligned} \mathrm{Var}(X_k) &= \mathcal{E}\{[I_k^2 - Q_k^2]^2\} - \mathcal{E}^2\{I_k^2 - Q_k^2\} \\ &= \mathcal{E}\{I_k^4 + Q_k^4 - 2I_k^2 Q_k^2\} - \mu_k^2 \end{aligned} \tag{A-2}$$

The covariance of X_k with X_{k+j} is

$$\mathrm{Cov}(X_k, X_{k+j}) = \mathcal{E}\{I_k^2 I_{k+1}^2\} + \mathcal{E}\{Q_k^2 Q_{k+j}^2\} - \mathcal{E}\{I_k^2 Q_{k+j}^2\} - \mathcal{E}\{I_{k+j}^2 Q_k^2\} - \mu_k \mu_{k+j} \tag{A-3}$$

When the loop is in lock, Eqs. (A-1) through (A-3) are evaluated with τ set to zero in Eqs. (A-10) through (A-18). Hence, the in-lock moments of X_k are given by

$$\mathrm{Var}(X_k) = \frac{A^4 T^4}{4} + 8A^2 T^2 \sigma_n^2 + 12\sigma_n^4 \tag{A-4}$$

and

$$\mathrm{Cov}(X_k, X_{k+1}) = -2A^2 T^2 \sigma_n^2 - 2\sigma_n^4 \tag{A-5}$$

and for $j \geq 2$, this can be shown to be

$$\mathrm{Cov}(X_k, X_{k+j}) = 0 \tag{A-6}$$

When the loop is out-of-lock, τ is modelled as a uniform random variable. Using this model for τ in Eqs. (A-10) through (A-18) and substituting the results into Eqs. (A-1) through (A-3) give the out-of-lock moments of X_k. Namely, (where the additional subscript o denotes out-of-lock),

$$\mathrm{Var}_o(I_k^2 - Q_k^2) = \frac{31A^4 T^4}{120} + \frac{41}{6} A^2 T^2 \sigma_n^2 + 12\sigma_n^4 \tag{A-7}$$

$$\mathrm{Cov}_o(X_k, X_{k+1}) = \frac{23A^4 T^4}{320} - \frac{23}{12} A^2 T^2 \sigma_n^2 - 2\sigma_n^4 \tag{A-8}$$

and for $j \geq 2$, this can be shown to be

$$\mathrm{Cov}_o(X_k, X_{k+j}) = \frac{A^4T^4}{12} \tag{A-9}$$

The following equations were used to compute the variance of X_k and covariance of X_k with X_{k+1}:

$$\mathcal{E}\{I_k^2\} = A^2(T^2 - 2T\mathcal{E}\{\tau\} + 2\mathcal{E}\{\tau^2\}) + 2\sigma_n^2 \tag{A-10}$$

$$\mathcal{E}\{Q_k^2\} = \begin{cases} A^2(\frac{T^2}{2}\mathcal{E}_1\{1\} + 2\mathcal{E}_1\{\tau^2\}) + 2\sigma_n^2\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^2(\frac{5T^2}{2}\mathcal{E}_2\{1\} - 4T\mathcal{E}_2\{\tau\} + 2\mathcal{E}_2\{\tau^2\}) + 2\sigma_n^2\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-11}$$

$$\mathcal{E}\{I_k^4\} = A^4(T^4 - 4T^3\mathcal{E}\{\tau\} + 12T^2\mathcal{E}\{\tau^2\} - 16T\mathcal{E}\{\tau^3\} + 8\mathcal{E}\{\tau^4\}) + A^2\sigma_n^2(12T^2 - 24T\mathcal{E}\{\tau\} + 24\mathcal{E}\{\tau^2\}) + 12\sigma_n^4 \tag{A-12}$$

$$\mathcal{E}\{Q_k^4\} = \begin{cases} A^4(\frac{T^4}{2}\mathcal{E}_1\{1\} + 8\mathcal{E}_1\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(6T^2\mathcal{E}_1\{1\} + 24\mathcal{E}_1\{\tau^2\}) + 12\sigma_n^4\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^4(\frac{17T^4}{2}\mathcal{E}_2\{1\} - 32T^3\mathcal{E}_2\{\tau\} + 48T^2\mathcal{E}_2\{\tau^2\} - 32T\mathcal{E}_2\{\tau^3\} + 8\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(30T^2\mathcal{E}_2\{1\} - 48T\mathcal{E}_2\{\tau\} + 24\mathcal{E}_2\{\tau^2\}) + 12\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-13}$$

$$\mathcal{E}\{I_k^2Q_k^2\} = \begin{cases} A^4(\frac{T^4}{2}\mathcal{E}_1\{1\} + 2T^2\mathcal{E}_1\{\tau^2\} - 8T\mathcal{E}_1\{\tau^3\} + 8\mathcal{E}_1\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(5T^2\mathcal{E}_1\{1\} - 8T\mathcal{E}_1\{\tau\} + 16\mathcal{E}_1\{\tau^2\}) + 6\sigma_n^4\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^4(\frac{5T^4}{2}\mathcal{E}_2\{1\} - 9T^3\mathcal{E}_2\{\tau\} + 15T^2\mathcal{E}_2\{\tau^2\} - 12T\mathcal{E}_2\{\tau^3\} + 4\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(7T^2\mathcal{E}_2\{1\} - 6T\mathcal{E}_2\{\tau\} + 4\mathcal{E}_2\{\tau^2\}) + 6\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-14}$$

$$\mathcal{E}\{I_k^2I_{k+1}^2\} = A^4(T^4 - 4T^3\mathcal{E}\{\tau\} + 8T^2\mathcal{E}\{\tau^2\} - 8T\mathcal{E}\{\tau^3\} + 4\mathcal{E}\{\tau^4\}) + A^2\sigma_n^2(4T^2 - 8T\mathcal{E}\{\tau\} + 8\mathcal{E}\{\tau^2\}) + 4\sigma_n^4 \tag{A-15}$$

$$\mathcal{E}\{Q_k^2Q_{k+1}^2\} = \begin{cases} A^4(\frac{T^4}{4}\mathcal{E}_1\{1\} + 2T^2\mathcal{E}_1\{\tau^2\} + 4\mathcal{E}_1\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(2T^2\mathcal{E}_1\{1\} + 8\mathcal{E}_1\{\tau^2\}) + 4\sigma_n^4\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^4(\frac{25T^4}{4}\mathcal{E}_2\{1\} - 20T^3\mathcal{E}_2\{\tau\} + 26T^2\mathcal{E}_2\{\tau^2\} - 16T\mathcal{E}_2\{\tau^3\} + 4\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(10T^2\mathcal{E}_2\{1\} - 16T\mathcal{E}_2\{\tau\} + 8\mathcal{E}_2\{\tau^2\}) + 4\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-16}$$

$$\mathcal{E}\{I_k^2 Q_{k+1}^2\} = \begin{cases} A^4(\frac{T^4}{2}\mathcal{E}_1\{1\} - T^3\mathcal{E}_1\{\tau\} + 3T^2\mathcal{E}_1\{\tau^2\} - 4T\mathcal{E}_1\{\tau^3\} + 4\mathcal{E}_1\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(3T^2\mathcal{E}_1\{1\} - 4T\mathcal{E}_1\{\tau\} + 8\mathcal{E}_1\{\tau^2\}) + 4\sigma_n^4\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^4(\frac{5T^4}{2}\mathcal{E}_2\{1\} - 9T^3\mathcal{E}_2\{\tau\} + 15T^2\mathcal{E}_2\{\tau^2\} - 12T\mathcal{E}_2\{\tau^3\} + 4\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(7T^2\mathcal{E}_2\{1\} - 12T\mathcal{E}_2\{\tau\} + 8\mathcal{E}_2\{\tau^2\}) + 4\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-17}$$

$$\mathcal{E}\{I_{k+1}^2 Q_k^2\} = \begin{cases} A^4(\frac{T^4}{2}\mathcal{E}_1\{1\} - T^3\mathcal{E}_1\{\tau\} + 3T^2\mathcal{E}_1\{\tau^2\} - 4T\mathcal{E}_1\{\tau^3\} + 4\mathcal{E}_1\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(5T^2\mathcal{E}_1\{1\} - 2T\mathcal{E}_1\{\tau\} + 4\mathcal{E}_1\{\tau^2\}) + 6\sigma_n^4\mathcal{E}_1\{1\} & 0 \le \tau < \frac{T}{2} \\ A^4(\frac{5T^4}{2}\mathcal{E}_2\{1\} - \frac{39T^3}{4}\mathcal{E}_2\{\tau\} + \frac{71T^2}{4}\mathcal{E}_2\{\tau^2\} - 15T\mathcal{E}_2\{\tau^3\} + 5\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(13T^2\mathcal{E}_2\{1\} - 24T\mathcal{E}_2\{\tau\} + 16\mathcal{E}_2\{\tau^2\}) + 6\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{2} \le \tau < T \end{cases} \tag{A-18}$$

where, in the above equations

$$\mathcal{E}\{f(\tau)\} \triangleq \int_0^T f(\tau)p(\tau)d\tau \tag{A-19}$$

$$\mathcal{E}_1\{f(\tau)\} \triangleq \int_0^{\frac{T}{2}} f(\tau)p(\tau)d\tau \tag{A-20}$$

$$\mathcal{E}_2\{f(\tau)\} \triangleq \int_{\frac{T}{2}}^T f(\tau)p(\tau)d\tau \tag{A-21}$$

where $p(\tau)$ is the probabilty density function of the variable τ.

Appendix B

Derivation of the Mean and Variance of the AVOD

Note that the calculations in this appendix incorportate the results of Appendix C. The inphase and quadraphase outputs are given, respectively, by Eqs. (13) and (14). The lock detector output $X_k = |I_k| - |Q_k|$. Let $\mu_k = \mathcal{E}\{|I_k| - |Q_k|\}$. Then,

$$\begin{aligned} \mathrm{Var}(X_k) &= \mathcal{E}\{[|I_k| - |Q_k|]^2\} - \mu_k^2 \\ &= \mathcal{E}\{I_k^2 + Q_k^2 - 2|I_k Q_k|\} - \mu_k^2 \end{aligned} \tag{B-1}$$

and

$$\begin{aligned} \mathrm{Cov}(X_k, X_{k+j}) &= \mathcal{E}\{[|I_k| - |Q_k| - \mu_k][|I_{k+j}| - |Q_{k+j}| - \mu_{k+j}]\} \\ &= \mathcal{E}\{|I_k I_{k+j}|\} + \mathcal{E}\{|Q_k Q_{k+j}|\} - \mathcal{E}\{|I_k Q_{k+j}|\} - \mathcal{E}\{|I_{k+j} Q_k|\} - \mu_k \mu_{k+j} \end{aligned} \tag{B-2}$$

The following equations were used to compute the variance of X_k and the covariance of X_k with X_{k+1}:

$$\mathcal{E}\{X_k\} = \begin{cases} \frac{1}{2}[\mathcal{E}_1\{|AT - 2A\tau + N|\} - \mathcal{E}_1\{|2A\tau + N|\}] & 0 \le \tau < \frac{T}{2} \\ \frac{1}{2}[\mathcal{E}_2\{|AT - 2A\tau + N|\} - \mathcal{E}_2\{|2AT - 2A\tau + N|\}] & \frac{T}{2} \le \tau < T \end{cases} \tag{B-3}$$

$$\mathcal{E}\{|I_k Q_k|\} = \begin{cases} \frac{1}{2}[\mathcal{E}_1\{|[AT + N_1(k) + N_2(k)][AT + N_2(k) + N_1(k+1)]|\} \\ \quad + \mathcal{E}_1\{|[AT - 2A\tau + N_1(k) + N_2(k)][-2A\tau + N_2(k) + N_1(k+1)]|\}] & 0 \le \tau < \frac{T}{2} \\ \frac{1}{4}[\mathcal{E}_2\{|[AT + N_1(k) + N_2(k)][AT + N_2(k) + N_1(k+1)]|\} \\ \quad + \mathcal{E}_2\{|[AT + N_1(k) + N_2(k)][2AT - 2A\tau + N_2(k) + N_1(k+1)]|\} \\ \quad + \mathcal{E}_2\{|[AT - 2A\tau + N_1(k) + N_2(k)][2AT - 2A\tau + N_2(k) + N_1(k+1)]|\} \\ \quad + \mathcal{E}_2\{|[AT - 2A\tau + N_1(k) + N_2(k)][AT + N_2(k) + N_1(k+1)]|\}] & \frac{T}{2} \le \tau < T \end{cases} \tag{B-4}$$

$$\begin{aligned} \mathcal{E}\{|I_k I_{k+1}|\} = \frac{1}{4}[&\mathcal{E}^2\{|AT + N_1(k) + N_2(k)|\} \\ &+ \mathcal{E}\{|[AT - 2A\tau + N_1(k) + N_2(k)][AT - 2A\tau + N_1(k+1) + N_2(k+1)]|\} \\ &+ 2\mathcal{E}\{|AT + N_1(k) + N_2(k)|\}\mathcal{E}\{|AT - 2A\tau + N_1(k+1) + N_2(k+1)|\}] \end{aligned} \tag{B-5}$$

$$\mathcal{E}\{|Q_kQ_{k+1}|\} = \begin{cases} \frac{1}{4}[\mathcal{E}_1^2\{|AT+N_1+N_2|\}+\mathcal{E}_1\{|[2A\tau+N_1(k+1)+N_2(k)] & \\ \quad\times[2A\tau+N_1(k+2)+N_2(k+1)]|\} & \\ \quad+2\mathcal{E}\{|AT+N_1+N_2|\}\mathcal{E}_1\{|2A\tau+N_1+N_2|\}] & 0\le\tau<\frac{T}{2} \\ \frac{1}{4}[\mathcal{E}_2^2\{|AT+N_1+N_2|\}+\mathcal{E}_2\{|[2AT-2A\tau+N_1(k+1)+N_2(k)] & \\ \quad\times[2AT-2A\tau+N_1(k+2)+N_2(k+1)]|\} & \\ \quad+2\mathcal{E}\{|AT+N_1+N_2|\}\mathcal{E}_2\{|2AT-2A\tau+N_1+N_2|\}] & \frac{T}{2}\le\tau<T \end{cases} \tag{B-6}$$

$$\mathcal{E}\{|I_kQ_{k+1}|\} = \begin{cases} \frac{1}{4}[\mathcal{E}^2\{|AT+N_1+N_2|\}+\mathcal{E}_1\{|[2A\tau+N_1(k+1)+N_2(k)] & \\ \quad\times[AT-2A\tau+N_1(k+2)+N_2(k+1)]|\} & \\ \quad+2\mathcal{E}\{|AT+N_1+N_2|\}\mathcal{E}_1\{|2A\tau+N_1+N_2|\}] & 0\le\tau<\frac{T}{2} \\ \frac{1}{4}[\mathcal{E}^2\{|AT+N_1+N_2|\}+\mathcal{E}_2\{|[AT-2A\tau+N_1(k+1)+N_2(k)] & \\ \quad\times[2AT-2A\tau+N_1(k+2)+N_2(k+1)]|\} & \\ \quad+2\mathcal{E}\{|AT+N_1+N_2|\}\mathcal{E}\{|AT-2A\tau+N_1+N_2|\}] & \frac{T}{2}\le\tau<T \end{cases} \tag{B-7}$$

$$\mathcal{E}\{|I_{k+1}Q_k|\} = \begin{cases} \frac{1}{4}[\mathcal{E}_1\{|[AT+N_1(k+1)+N_2(k+1)][AT+N_1(k+1)+N_2(k)]|\} & \\ \quad+\mathcal{E}_1\{|[AT-2A\tau+N_1(k+1)+N_2(k+1)][AT+N_1(k+1)+N_2(k)]|\} & \\ \quad+\mathcal{E}_1\{|[AT-2A\tau+N_1(k+1)+N_2(k+1)][2A\tau+N_1(k+1)+N_2(k)]|\} & \\ \quad+\mathcal{E}_1\{|[AT+N_1(k+1)+N_2(k+1)][2A\tau+N_1(k+1)+N_2(k)]|\}] & 0\le\tau<\frac{T}{2} \\ \frac{1}{2}[\mathcal{E}_2\{|[AT-2A\tau+N_1(k+1)+N_2(k+1)][2AT-2A\tau+N_1(k+1)+N_2(k)]|\} & \\ \quad+\mathcal{E}_2\{|[AT+N_1(k+1)+N_2(k+1)][AT+N_1(k+1)+N_2(k)]|\}] & \frac{T}{2}\le\tau<T \end{cases} \tag{B-8}$$

where $\mathcal{E}_1$ and $\mathcal{E}_2$ are defined in Appendix A. The following functions were defined to obtain the results in Subsection III.A of the main text.

$$F_1(\eta_s) \triangleq \mathcal{E}\{|(1+cn_1+cn_2)(1+cn_2+cn_3)|\}$$

$$F_2(\eta_s) \triangleq \mathcal{E}\{|(1+cn_1+cn_2)(cn_2+cn_3)|\}$$

$$F_3(\eta_s) \triangleq \mathcal{E}\{|(1-2u+cn_1+cn_2)(1-2u+cn_3+cn_4)|\}$$

$$G_1(\eta_s) \triangleq \mathcal{E}_1\{|(1-2u+cn_1+cn_2)(2u+cn_2+cn_3)|\}$$

$$G_2(\eta_s) \triangleq \mathcal{E}_1\{|(1-2u+cn_1+cn_2)(1+cn_2+cn_3)|\}$$

$$G_3(\eta_s) \triangleq \mathcal{E}_1\{|(2u + cn_1 + cn_2)(2u + cn_3 + cn_4)|\}$$

$$G_4(\eta_s) \triangleq \mathcal{E}_1\{|(2u + cn_1 + cn_2)(1 - 2u + cn_3 + cn_4)|\}$$

$$H_1(\eta_s) \triangleq \mathcal{E}_2\{|(1 + cn_1 + cn_2)(2 - 2u + cn_2 + cn_3)|\}$$

$$H_2(\eta_s) \triangleq \mathcal{E}\{|(1 - 2u + cn_1 + cn_2)(2 - 2u + cn_2 + cn_3)|\}$$

$$H_3(\eta_s) \triangleq \mathcal{E}\{|(1 - 2u + cn_1 + cn_2)(1 + cn_2 + cn_3)|\}$$

$$H_4(\eta_s) \triangleq \mathcal{E}\{|(2 - 2u + cn_1 + cn_2)(2 - 2u + cn_3 + cn_4)|\}$$

$$H_5(\eta_s) \triangleq \mathcal{E}\{|(1 - 2u + cn_1 + cn_2)(2 - 2u + cn_3 + cn_4)|\} \qquad \text{(B-9)}$$

where $c \triangleq 1/(2\sqrt{\eta_s})$, the n_i's are normal independent random variables with zero mean and unit variance, and u is a uniform random variable in the range $[0, 1]$. In Fig. B-1, one plots these functions versus η_s. These functions have been computed as follows: In the F functions, the expectation with respect to u is carried over the entire region [0, 1], while in G and H functions, the expectation is carried over [0, 1/2] and [1/2, 1], respectively.

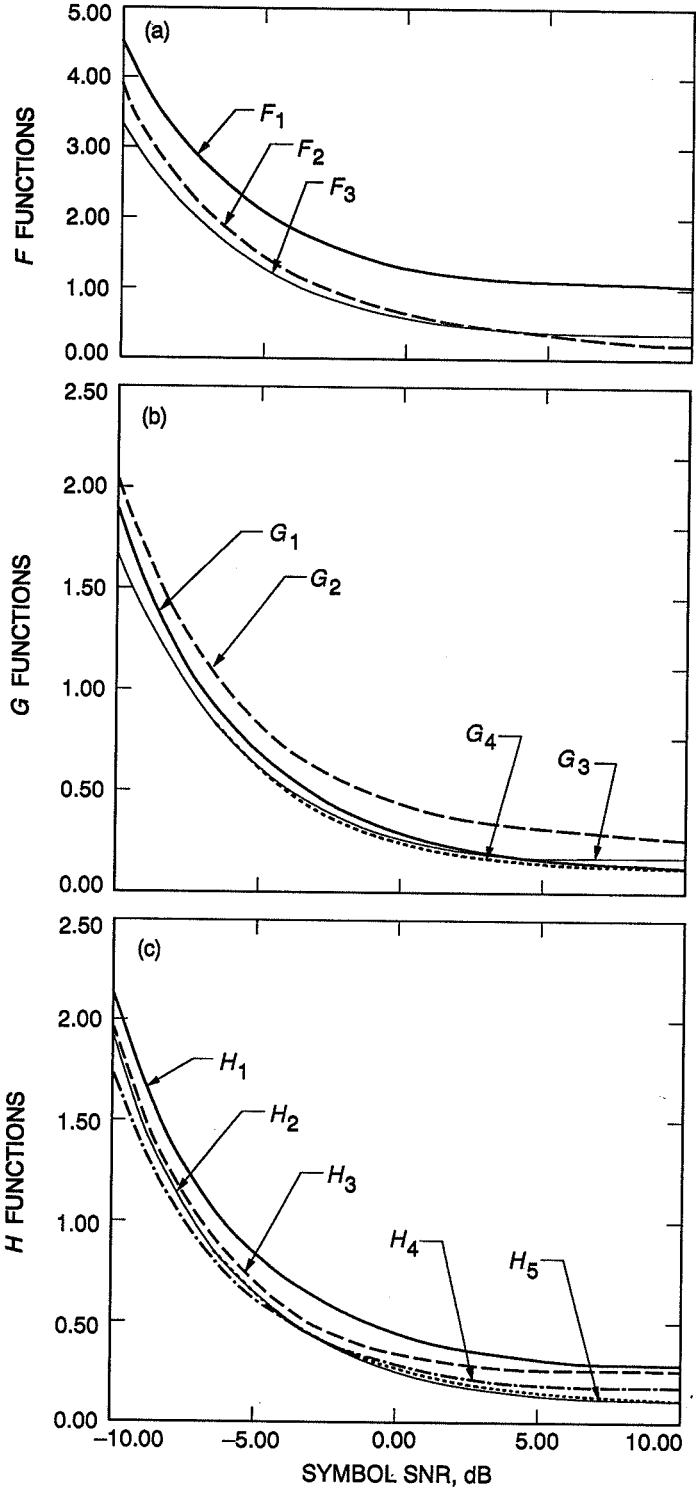


Fig. B-1. Function versus symbol SNR for: (a) *F* functions, (b) *G* functions, and (c) *H* functions.

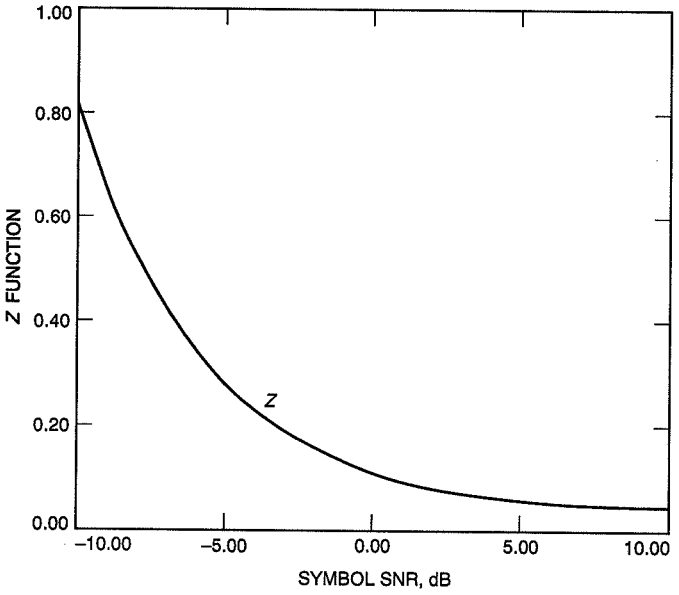


Fig. B-2. *Z* function versus symbol SNR.

Appendix C

Derivation of the Mean for the Random Variable $|n + b\tau + c|$

Let n be a normal random variable with zero mean and variance σ^2, τ a uniform random variable over $(0, T)$, and c any constant. Then,

$$\mathcal{E}\{|n+c|\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} |n+c| \exp\left(-\frac{n^2}{2\sigma^2}\right) dn \tag{C-1}$$

The above integral can be easily evaluated by breaking it into two integrals over the two regions $(-\infty, -c)$ and $(-c, \infty)$ to obtain

$$\mathcal{E}\{|n+c|\} = \sqrt{\frac{2\sigma^2}{\pi}} \exp\left(-\frac{c^2}{2\sigma^2}\right) + c \operatorname{erf}\left(\frac{c}{\sqrt{2\sigma^2}}\right) \tag{C-2}$$

For a fixed τ, one can write

$$\mathcal{E}\{|n+b\tau+c|\} = \sqrt{\frac{2\sigma^2}{\pi}} \exp\left(-\frac{(b\tau+c)^2}{2\sigma^2}\right) + (b\tau+c)\operatorname{erf}\left(\frac{b\tau+c}{\sqrt{2\sigma^2}}\right) \tag{C-3}$$

Unconditioning over τ yields

$$\mathcal{E}_1\{|n+b\tau+c|\} = \frac{1}{T}\int_0^{\frac{T}{2}} \left[\sqrt{\frac{2\sigma^2}{\pi}} \exp\left(-\frac{(b\tau+c)^2}{2\sigma^2}\right) + (b\tau+c)\operatorname{erf}\left(\frac{b\tau+c}{\sqrt{2\sigma^2}}\right)\right] d\tau \tag{C-4}$$

which, after integration by parts, leads to

$$\mathcal{E}_1\{|n+b\tau+c|\} = \frac{\sigma^2}{2bT}\left[\operatorname{erf}\left(\frac{\frac{bT}{2}+c}{\sqrt{2\sigma^2}}\right) - \operatorname{erf}\left(\frac{c}{\sqrt{2\sigma^2}}\right)\right] + \frac{1}{2bT}\left[\left(\frac{bT}{2}+c\right)^2 \operatorname{erf}\left(\frac{\frac{bT}{2}+c}{\sqrt{2\sigma^2}}\right) - c^2\operatorname{erf}\left(\frac{c}{\sqrt{2\sigma^2}}\right)\right]$$

$$+ \sqrt{\frac{\sigma^2}{2\pi}}\frac{1}{bT}\left[\left(\frac{bT}{2}+c\right)\exp\left(\frac{-(\frac{bT}{2}+c)^2}{2\sigma^2}\right) - c\exp\left(\frac{-c^2}{2\sigma^2}\right)\right] \tag{C-5}$$

By applying the above expression, one gets

$$\mathcal{E}_1\{|AT - 2A\tau + N_1 + N_2|\} = \sqrt{N_0 T}\left[\frac{1}{8\sqrt{\eta_s}}\operatorname{erf}(\sqrt{\eta_s}) + \frac{\sqrt{\eta_s}}{4}\operatorname{erf}(\sqrt{\eta_s}) + \frac{1}{4\sqrt{\pi}}\exp(-\eta_s)\right] \tag{C-6}$$

Also, by simple manipulation, it can be shown that

$$\begin{aligned}\mathcal{E}_1\{|AT - 2A\tau + N_1 + N_2|\} &= \mathcal{E}_2\{|AT - 2A\tau + N_1 + N_2|\} \\ &= \mathcal{E}_1\{|2A\tau + N_1 + N_2|\} \\ &= \mathcal{E}_2\{|2AT - 2A\tau + N_1 + N_2|\}\end{aligned} \tag{C-7}$$

Appendix D

Derivation of the Mean and Variance of the SQNOD

The samples I_k and Q_k are given by Eqs. (33) and (34). The moments of X_k are given by Eqs. (A-1) through (A-3). Using Eqs. (D-7) through (D-10) with $\tau = 0$ and Eqs. (A-2) and (A-3) yields the in-lock variance. Namely,

$$\mathrm{Var}(X_k) = \frac{A^4T^4}{64} + \frac{3}{2}A^2T^2\sigma_n^2 + 4\sigma_n^4 \tag{D-1}$$

The covariance can be shown to be, for $j \geq 1$,

$$\mathrm{Cov}(X_k, X_{k+j}) = 0 \tag{D-2}$$

Similarly, with τ modelled as uniform over $[0,\ T/2]$ in Eqs. (D-7) through (D-10), the out-of-lock moments are found. Hence,

$$\mathrm{Var}_o(X_k) = \frac{A^4T^4}{60} + \frac{5}{3}A^2T^2\sigma_n^2 + 4\sigma_n^4 \tag{D-3}$$

The out-of-lock covariance can be shown to be, for $j \geq 1$,

$$\mathrm{Cov}_o(X_k, X_{k+j}) = \frac{A^4T^4}{120} \tag{D-4}$$

The following equations are used in computing the variance of X_k:

$$\mathcal{E}\{I_k^2\} = \begin{cases} (\frac{A^2T^2}{4} + \sigma_n^2)\mathcal{E}_1\{1\} & 0 \leq \tau < \frac{T}{4} \\ A^2(\frac{5T^2}{8}\mathcal{E}_2\{1\} - 2T\mathcal{E}_2\{\tau\} + 2\mathcal{E}_2\{\tau^2\}) + \sigma_n^2\mathcal{E}_2\{1\} & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \tag{D-5}$$

$$\mathcal{E}\{Q_k^2\} = \begin{cases} A^2(\frac{T^2}{8}\mathcal{E}_1\{1\} + 2\mathcal{E}_1\{\tau^2\}) + \sigma_n^2\mathcal{E}_1\{1\} & 0 \leq \tau < \frac{T}{4} \\ (\frac{A^2T^2}{4} + \sigma_n^2)\mathcal{E}_2\{1\} & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \tag{D-6}$$

$$\mathcal{E}\{I_k^4\} = \begin{cases} (\frac{A^4T^4}{16} + \frac{3}{2}A^2T^2\sigma_n^2 + 3\sigma_n^4)\mathcal{E}_1\{1\} & 0 \leq \tau < \frac{T}{4} \\ A^4(\frac{17T^4}{32}\mathcal{E}_2\{1\} - 4T^3\mathcal{E}_2\{\tau\} + 12T^2\mathcal{E}_2\{\tau^2\} - 16T\mathcal{E}_2\{\tau^3\} + 8\mathcal{E}_2\{\tau^4\}) & \\ \quad +A^2\sigma_n^2(\frac{15}{4}T^2\mathcal{E}_2\{1\} - 12T\mathcal{E}_2\{\tau\} + 12\mathcal{E}_2\{\tau^2\}) + 3\sigma_n^4\mathcal{E}_2\{1\} & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \tag{D-7}$$

$$\mathcal{E}\{Q_k^4\} = \begin{cases} A^4(\frac{T^4}{32}\mathcal{E}_1\{1\} + 8\mathcal{E}_1\{\tau^4\}) + A^2\sigma_n^2(\frac{3}{4}T^2\mathcal{E}_1\{1\} + 12\mathcal{E}_1\{\tau^2\}) + 3\sigma_n^4\mathcal{E}_1\{1\} & 0 \leq \tau < \frac{T}{4} \\ (\frac{A^4T^4}{16} + \frac{3}{2}A^2T^2\sigma_n^2 + 3\sigma_n^4)\mathcal{E}_2\{1\} & \frac{T}{4} \leq \tau < \frac{T}{2} \end{cases} \tag{D-8}$$

Appendix E

Derivation of the Mean and Variance of the AVNOD

By following the same procedure as in Appendix B and by using

$$\mathcal{E}\{|I_k|\} = \begin{cases} \mathcal{E}_1\{|\frac{AT}{2} + N|\} & 0 \le \tau < \frac{T}{4} \\ \frac{1}{2}[\mathcal{E}_2\{|\frac{AT}{2} + N|\} + \mathcal{E}_2\{|AT - 2A\tau + N|\}] & \frac{T}{4} \le \tau < \frac{T}{2} \end{cases} \tag{E-1}$$

$$\mathcal{E}\{|Q_k|\} = \begin{cases} \frac{1}{2}[\mathcal{E}_1\{|\frac{AT}{2} + N|\} + \mathcal{E}_1\{|2A\tau + N|\}] & 0 \le \tau < \frac{T}{4} \\ \mathcal{E}_2\{|\frac{AT}{2} + N|\} & \frac{T}{4} \le \tau < \frac{T}{2} \end{cases} \tag{E-2}$$

and

$$\mathrm{Var}(|I_k| - |Q_k|) = \begin{cases} \frac{3A^2T^2}{8}\mathcal{E}_1\{1\} + 2A^2\mathcal{E}_1\{\tau^2\} + 2\sigma_n^2\mathcal{E}_1\{1\} \\ \quad -\frac{5}{4}\mathcal{E}_1^2\{|\frac{AT}{2} + N|\} - \frac{1}{4}\mathcal{E}_1^2\{|2A\tau + N|\} \\ \quad -\frac{1}{2}\mathcal{E}\{|\frac{AT}{2} + N|\}\mathcal{E}_1\{|2A\tau + N|\} & 0 \le \tau < \frac{T}{4} \\ \frac{7A^2T^2}{8}\mathcal{E}_2\{1\} - 2A^2T\mathcal{E}_2\{\tau\} + 2A^2\mathcal{E}_2\{\tau^2\} + 2\sigma_n^2\mathcal{E}_2\{1\} \\ \quad -\frac{5}{4}\mathcal{E}_2^2\{|\frac{AT}{2} + N|\} - \frac{1}{4}\mathcal{E}_2^2\{|AT - 2A\tau + N|\} \\ \quad -\frac{1}{2}\mathcal{E}\{|\frac{AT}{2} + N|\}\mathcal{E}_2\{|AT - 2A\tau + N|\} & \frac{T}{4} \le \tau < \frac{T}{2} \end{cases} \tag{E-3}$$

one obtains Eqs. (40) through (45) after using the results of Appendix C and lengthy manipulations. The function Z in Eq. (45) is defined as

$$Z(\eta_s) \triangleq \mathcal{E}_1\{|(u + cn_1)(u + cn_2)|\} \tag{E-4}$$

where $c \triangleq 1/(2\sqrt{\eta_s})$, the n_i's are normal independent random variables with zero mean and unit variance, and u is a uniform random variable in the range (0, 1). In Z, the expectation with respect to u is over the range [0, 1/4] and the function is plotted in Fig. B-2 versus the symbol SNR (η_s).

Appendix F

Derivation of the Mean and Variance of the SPED

The samples I_k and Q_k are given by Eqs. (47) and (48). The output of the lock detector is $x_k = I_k Q_k$. It is straightforward to show that

$$\mathcal{E}\{X_k\} = \frac{A^2T^2}{4} - \frac{A^2T}{2}\mathcal{E}\{\tau\} \tag{F-1}$$

and

$$\mathcal{E}\{X_k^2\} = A^4\left(\frac{T^4}{16} - \frac{T^3\mathcal{E}\{\tau\}}{4} + \frac{T^2\mathcal{E}\{\tau^2\}}{2}\right) + \sigma_n^4 + A^2\sigma_n^2\left(\frac{T^2}{2} - T\mathcal{E}\{\tau\} + 2\mathcal{E}\{\tau^2\}\right) \tag{F-2}$$

$$\mathrm{Cov}(X_k, X_{k+j}) = \begin{cases} \frac{A^4T^2}{4}\mathrm{Var}(\tau) & \tau \sim u(0, \frac{T}{2}) \text{ and } j \geq 1 \\ 0 & \tau = 0 \end{cases} \tag{F-3}$$

Equations (49) and (50) follow after letting $\tau = 0$ in Eqs. (F-1) through (F-3). Equations (51) and (52) follow by letting τ be uniform over $[0,\ T/2]$.

Design and Status of the RF-Digitizer Integrated Circuit

B. Rayhrer, B. Lam, L. E. Young, J. M. Srinivasan, and J. B. Thomas
Tracking Systems and Applications Section

An integrated circuit currently under development samples a bandpass-limited signal at a radio frequency (RF) in quadrature and then performs a simple sum-and-dump operation in order to filter and lower the rate of the samples. Downconversion to baseband is carried out by the sampling step itself through the aliasing effect of an appropriately selected "subharmonic" sampling frequency. Two complete RF digitizer circuits with these functions will be implemented with analog and digital elements on one GaAs substrate. An input signal, with a carrier frequency as high as 8 GHz, can be sampled at a rate as high as 600 Msamples/sec for each quadrature component. The initial version of the chip will sign-sample (1-bit) the input RF signal. The chip will contain a synthesizer to generate a sample frequency that is a selectable integer multiple of an input reference frequency. In addition to the usual advantages of compactness and reliability associated with integrated circuits, the single chip will replace several steps required by standard analog downconversion. Furthermore, when a very high initial sample rate is selected, the presampling analog filters can be given very large bandwidths, thereby greatly reducing phase and delay instabilities typically introduced by such filters, as well as phase and delay variations due to Doppler changes.

I. Introduction

Conventional digital systems for processing RF signals sample the signal after it has been downconverted from RF to lower frequencies by means of analog mixers and filters. Analog downconversion and its required steps and components can be eliminated by simply sampling the signal at RF, after appropriate bandpass filtering, at a sample rate that is commensurate with the signal carrier. This process, sometimes called subharmonic sampling, aliases the signal directly to baseband.

In addition to the elimination of a number of analog components and the usual advantages of compactness and reliability provided by digital processing, another advantage can be realized. If the sample rate can be made very large, the phase and delay instabilities due to the analog filters preceding sampling can be made very small by increasing the bandwidths of the filters. For example, with a presampling bandwidth of 600 MHz, the filter delay can be reduced to the order of two nanoseconds in magnitude and a fraction of a nanosecond in instability. In addition, a larger analog bandwidth can greatly reduce phase and delay sensitivity to Doppler variation in the carrier.

This article outlines the design and status of a GaAs integrated circuit (IC) that is designed to sample an RF-bandpass signal at a rate as high as 1200 Msamples/sec (600 Msamples/sec per quadrature component) when the

carrier frequency is as high as 8 GHz. The IC, which will be referred to as the RF digitizer (RFD), will contain a synthesizer that will generate a sample frequency that is a selectable integer multiple of an input reference frequency. This capability has been included in the IC in order to satisfy the stringent stability requirements set by RF sampling. A digital filter consisting of a simple sum-and-dump operation has been included in the IC in order to allow the user the option of reducing an initial high sample rate to a lower value that can be easily processed by subsequent digital circuitry.

Initial use of the RFD chip is planned for the Global Positioning Satellite (GPS) receiver portion of the media calibration subassembly at the Deep Space Network (DSN) antennas. The current chip design contains a pair of RF digitizers with matched synthesizers and samplers. This design greatly improves the differential stability in phase and delay between the two GPS frequency bands (L1 and L2), thereby improving ionospheric calibration.

II. Subharmonic Sampling

The technique of subharmonic sampling is based on the aliasing that occurs when a signal is sampled at a given rate. If a signal component at frequency f_r is sampled at the rate f_r/n, where n is an integer, that component will be aliased to zero frequency. Signal components with nearby frequencies are aliased to baseband as though they had been downconverted using a mixing signal with frequency f_r. A more sophisticated downconversion process can be carried out with "half-subharmonic" sampling [1,2,3], in which the sample rate becomes $f_s = (4f_r)/(2n+1)$, where n is a non-negative integer and f_r is again the frequency component to be downconverted to zero frequency. The advantage of half-subharmonic sampling is that a single sampler at RF, driven by one sample clock, directly produces both quadrature components at baseband with exact quadrature separation. More specifically stated, the sampling process directly produces interleaved cosine (C) and sine (S) samples at baseband in the sequence C, S, −C, −S, C, S, −C, −S, ···. The indicated sign reversals on alternate C samples (and S samples) are removed by subsequent IC logic. Because the half-subharmonic approach very simply produces both C and S baseband samples with very exact quadrature separation, this approach has been used on the IC.

III. Functional Description

A functional block diagram of the RFD is shown in Fig. 1. It is assumed that the input RF signal, which can have a carrier as high as 8 GHz, is filtered to satisfy the Nyquist criterion as applied to bandpass sampling. That is, if the initial sample rate is f_s (0.5 f_s on C and S separately), the sampled signal should have a bandwidth equal to or less than 0.5 f_s. For example, if the initial sample rate is set at the maximum value of 1200 Msamples/sec (Ms/s), then the separate C and S rates will be 600 Ms/s and the maximum input bandwidth can be as high as 600 MHz. The selectable sample rate can range between 4 Ms/s and 1200 Ms/s. Although the initial version of the RFD sign samples (1-bit) the input signal, a second-generation design is planned with 6-bit sampling.

As explained above, half-subharmonic sampling produces interleaved C and S samples with alternating sign reversals. In the IC, the C and S samples are demultiplexed into separate streams and a sign correction is applied to alternate samples of each stream. Each stream is then subjected to a sum-and-dump operation in order to filter the signal and lower the data rate. The sum interval is user selectable, with suggested values of $N_s = 1, 2, 4, 6, 8, \cdots$, or 30. After summing, the 5-bit sum values can be requantized to three levels (−1, 0, and +1), at the user's option. The resulting values and associated clock signal are then output from the IC at a rate of $(0.5f_s)/N_s$ for each quadrature component.

IV. Detailed Description

Two RFD's, each incorporating both analog and digital components, will be placed on one GaAs substrate. GaAs was chosen since it is expected to satisfy the stringent speed requirement necessitated by RF sampling. The most critical components of the IC are the frequency synthesizer, an ultra-fast sample-and-hold circuit, and the analog comparator for analog-to-digital (A/D) conversion. The IC will have a complexity equivalent to about 2000 gates, including two matched RFD's, and consume about 10 W. The die will be about 4 mm × 4 mm in size, and will be packaged in a 149-pin grid array.

The sample-rate "clock," which is generated on-chip by a phase-locked loop synthesizer, is a selectable integer multiple of an input reference frequency between 1 and 200 MHz. The user-selectable multiplier value is equal to $4N_m$, where $4N_m = 4, 8, 12, \cdots, 1024$. Jitter introduced by the synthesizer is expected to be less than 10 psec.

The sampler circuit will consist of two samplers (primary and secondary). The primary sampler, a Schottky diode bridge, satisfies the high-speed requirement and will be constructed with the shortest possible sampling window (about 20 psec). Placed between the primary and

secondary samplers is an amplifier with a gain of 10 dB and a settling time less than 400 psec. This amplifier increases the signal strength to a level (1 V) that is adequate for the secondary sampler and the A/D converter. The purpose of the secondary sampler is to hold the output voltage of the primary sampler at a stable level for an entire clock period ($1/f_s$), thereby allowing the subsequent A/D conversion circuit ample time to settle. The secondary sampler consists of a diode circuit similar to the primary sampler.

The A/D converter translates the signal output by the secondary sampler to a 1-bit signal representing the sign. It consists of a voltage comparator, with a reference voltage of zero and a conversion time of less than 800 psec, operating at a rate of up to 1.2 GHz.

After demultiplexing and sign correction, each sample stream is subjected to a sum-and-dump operation to lower the data rate. To achieve the necessary speed, the sum operation is divided into two successive operations: a two-sample sum (1-bit adder) followed by an N-sample sum (5-bit adder). Note that this sum-and-dump design leads to total sum intervals that are even integers. The sum interval is user selectable with recommended values of $N_s = 2, 4, 6, 8, \cdots$, or 30. (Values as high as 510 can be selected, but adder overflow may result.) The user can also elect to bypass the sum and output the initial 1-bit samples at the full initial rate.

The sum values produced by the sum-and-dump operation are effectively 5-bit integers that exit the sum operation in parallel on 5 output pins. (Since the least significant bit, LSB, is always 0, no pin is provided for that bit.) The full, allowable range for the output integers is -30 to $+30$, including the phantom LSB.

To allow the user to significantly decrease the bit output rate, the user is given the option of commanding the IC to requantize the sum-and-dump integers from 5 bits to 2 bits, where the two-bit option represents -1, 0, and $+1$. In this requantization, integer discriminator levels are set so that the sum values of -1, 0, and $+1$ are converted to 0, the values of -2 and lower to -1, and the values of $+2$ or greater to $+1$. These discriminator levels minimize the signal-to-noise ratio (SNR) loss caused by the requantization when the sum length is about 20, correlation between initial samples is small, and the single-sample SNR is much less than one.

Data output by the IC can be passed in parallel form or converted to serial form, at the option of the user. The serial output can be made compatible with single-ended emitter-coupled logic.

V. Status and Plans

Most of the design, simulation, and layout for the IC have been completed and the IC is in the review process. To obtain estimates for IC characteristics (such as parasitic capacitance) that are needed for critical additional simulations of chip design, a test fabrication of a partial chip was submitted during the first quarter of calendar 1991. This partial chip is to be tested during the second quarter of 1991. The fabrication run for the complete 1-bit RFD will be submitted during the third quarter. After low-speed tests have confirmed the basic functionality of the IC, the high-frequency limits will be determined. Detailed measurements of SNR, phase, and delay performance will be carried out using the digital baseband processor of the Rogue GPS receiver [4].

The first implementation of the RFD is planned for the Rogue receiver installed at DSN sites to provide both ionospheric calibrations and precise orbit determination for low Earth satellites. The RFD will provide greater relative stability between signals measured at the two GPS frequencies. The flexible design of the RFD will also allow its use in other applications of high-speed digital processing, such as digital spacecraft transponders and high-accuracy radiometric receivers.

Further improvements are possible for future RFD chip implementations. A laser diode driver would allow the use of a fiber optic cable for transmission of time-tagged digital data. Improving the RF sampling from 1-bit to 6-bit quantization would make the RFD more suitable for low SNR applications such as future very long baseline interferometry receivers and advanced telemetry receivers.

References

[1] J. L. Brown, Jr., "On Quadrature Sampling of Bandpass Signals," *IEEE Trans. Aero. and Electr. Systems*, vol. AES-15, no. 3, pp. 366–371, 1979.

[2] O. D. Grace and S. P. Pitt, "Quadrature Sampling of High Frequency Waveforms," *J. Acoust. Soc. Amer.*, vol. 44, pp. 1453–1454, 1968.

[3] M. C. Jackson and P. Matthewson, "Digital Processing of Bandpass Signals," *GEC Journal of Research*, vol. 4, no. 1, pp. 32–41, 1986.

[4] J. B. Thomas, *Functional Description of Signal Processing in the Rogue GPS Receiver*, JPL Publication 88-15, Jet Propulsion Laboratory, Pasadena, California, June 1988.

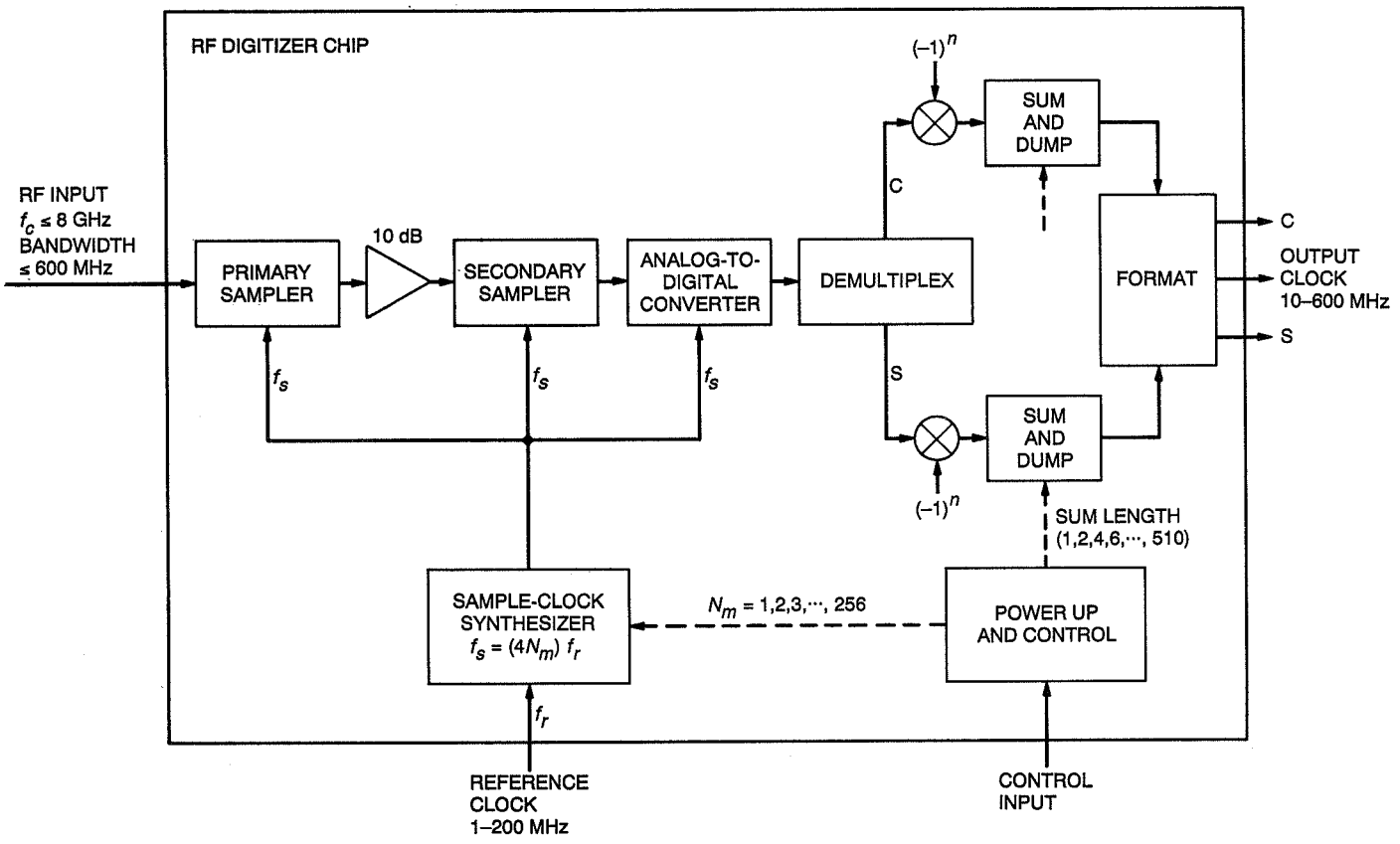


Fig. 1. Functional block diagram of the radio frequency digitizer.

Channel Assignments and Array Gain Bounds for the Ka-Band Array Feed Compensation System

V. A. Vilnrotter
Communications Systems Research Section

The performance of a real-time digital combining system for use with array feeds has been considered in previous articles. The purpose of the combining operation is to recover signal-to-noise ratio (SNR) losses due to antenna deformations and atmospheric effects. Previously, arbitrary signal powers and noise variances were assumed, but no attempt was made to match the receiver channels to the available signal powers. Here it is shown that for any signal power and noise variance distribution, a "best" channel assignment exists that maximizes the combined SNR in the limit of vanishingly small combining losses. This limit can be approached in practice by observing sufficiently many samples. Specific signal power and noise variance distributions are considered, and it is shown that even relatively "noisy" channels can be used effectively to recover SNR losses resulting from signals diverted out of a "high-quality" channel by antenna deformations.

I. Introduction

The potential benefits of array feed combining for recovering losses due to mechanical antenna distortions at high frequencies (32 GHz or higher) have been described in [1,2]. A conceptual block diagram of the real-time combining system considered here is shown in Fig. 1, where the dashed curve near the primary reflector represents mechanical distortions, while near the feed array it indicates the spreading of the signal field distribution induced by reflector distortions. The feedhorn outputs are downconverted to baseband, sampled at the Nyquist rate, and the samples are processed digitally to obtain the combined complex baseband samples. The objective of the combining operation is to maximize the signal-to-noise ratio (SNR) of the combined samples, thus recovering some of the losses induced by the distorted antenna. The seven-element feed array geometry is of particular interest, as it has been shown in [3] that most of the losses incurred by large Deep Space Network (DSN) antennas can be recovered by seven elements arranged in a maximally compact pattern, as in Fig. 1. However, the analysis and results apply for arbitrary K-element arrays, regardless of the array geometry.

A functional block diagram of the real-time compensation system is given in Fig. 2. The observables consist of K complex sample-streams, corrupted by independent complex noise samples in each channel. The corrupted samples are input to a parameter estimator, which estimates the complex weights that should be applied to each stream in order to maximize the SNR of the combined samples.

II. System Description

An exact expression for the signal-to-noise ratio ρ_{ML} of the combined sample-stream has been derived in [2] for arbitrary SNR in each channel, assuming use of a "maximum-likelihood" (ML) parameter estimator to obtain the combining weights. The combined SNR was found to depend on the number of channels K, the total number of observed samples L, the ratio of data to estimation stream bandwidths η, the modulation index δ, the sum of the channel SNRs, and the sum of squares of the channel SNRs. The complex weights that maximize the combined SNR are

$$\tilde{w}_k = \frac{\tilde{V}_k^*}{2\sigma_k^2} \tag{1a}$$

in which case the combined SNR is given by

$$\rho = \sum_{k=1}^{K} |\tilde{V}_k|^2/2\sigma_k^2 \tag{1b}$$

where $\tilde{V}_k$ is the complex voltage due to signal, and $2\sigma_k^2$ is the variance of the complex noise in the kth channel. However, in the presence of noise one must rely on estimates of the combining weights, with inadvertent errors. If maximum-likelihood estimates are employed, the combined SNR becomes

$$\rho_{ML} = \rho\left\{1+\frac{1}{L-3}+\frac{L-2}{\eta L(L-3)\cos^2\delta}\left(1+\frac{K}{\rho}\right)+\frac{1}{(L-3)\rho}\sum_{k=1}^{K}\frac{|\tilde{V}_k|^4}{4\sigma_k^4}\right\}^{-1} \tag{2}$$

Note that the combined SNR with maximum-likelihood weights is always less than ρ, approaching that limit as L approaches infinity. This is reasonable, since the maximum-likelihood weight estimates approach the true weights as the number of observed samples approaches infinity (the estimator is "consistent"). Note that for $L < 4$, this expression is not defined, since not enough samples have been observed to make all of the required estimates.

It is reasonable to define the "combining loss" γ as the inverse of the ratio of the actual SNR to its limiting value:

$$\gamma \triangleq \frac{\rho}{\rho_{ML}} \tag{3}$$

All losses due to imperfect weight estimates can be attributed to γ, where its inverse γ^{-1} is simply the variance of the combined signal normalized by ρ. This quantity can be decomposed into loss components due to signal γ_s and noise γ_n,

$$\gamma = \gamma_s + \gamma_n \tag{4a}$$

where

$$\gamma_s = \frac{L-2}{\eta L(L-3)\cos^2\delta}+\frac{1}{(L-3)\rho}\sum_{k=1}^{K}\frac{|\tilde{V}_k|^4}{4\sigma_k^4} \tag{4b}$$

$$\gamma_n = 1+\frac{1}{L-3}+\frac{K(L-2)}{\eta L(L-3)\rho\cos^2\delta} \tag{4c}$$

The behavior of the combining loss as a function of L is examined in Section V.

III. Receiver Channel Assignment

Suppose the total signal power captured by the array is P_T watts, with the kth horn contributing P_k watts to the total, $P_k = |V_k|^2$, so that

$$P_T = \sum_{k=1}^{K} P_k \tag{5}$$

Assume the ordering $P_1 \geq P_2 \geq \cdots \geq P_K$, so that horn number 1 contributes the greatest signal component, horn number 2 the second greatest, and so on. For want of a better term, this can be called a "standard ordering" of the array feeds (in case of equalities, ordering becomes irrelevant). Suppose that each receiver channel adds independent noise components to the signal, with the kth channel contributing variance $2\sigma_k^2$. If receiver channels could be assigned to array elements in any order, how should the receivers be assigned to the feeds in order to maximize the SNR of the combined signal ρ_{ML}? To answer this question, one first maximizes the ideal SNR ρ.

A. Channel Assignment to Maximize ρ

Index the receiver channels according to the variance of the noise in that channel, with the least noisy channel called channel number 1, the second "quietest" channel number 2, etc., so that $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_K^2$. Assigning the receiver channels to the array feeds according to this indexing scheme leads to the sum ρ, which is called "standard channel assignment" to be consistent with the above terminology. Then

$$2\rho = \sum_{k=1}^{K} P_K/\sigma_K^2 = \frac{P_1}{\sigma_1^2} + \frac{P_2}{\sigma_2^2} + \cdots + \frac{P_K}{\sigma_K^2}$$

$$= \left(\sum_{\substack{k=1 \\ k\neq i \\ k\neq j}}^{K} \frac{P_k}{\sigma_k^2} \right) + \left(\frac{P_i}{\sigma_i^2} + \frac{P_j}{\sigma_j^2} \right) \qquad (6)$$

Here the last equality focuses on the ith and jth channels. With $i < j$, let

$$\sigma_j^2 = \sigma_i^2 + \Delta_{ij}\ , \qquad \Delta_{ij} \geq 0 \qquad (7a)$$

$$P_i = P_j + \delta_{ij}\ , \qquad \delta_{ij} \geq 0 \qquad (7b)$$

hence, the sum of the ith and jth channel SNRs may be written as

$$\left(\frac{P_i}{\sigma_i^2} + \frac{P_j}{\sigma_j^2} \right) = \frac{P_i}{\sigma_i^2} + \frac{(P_i - \delta_{ij})}{(\sigma_i^2 + \Delta_{ij})} \qquad (8a)$$

Exchange the ith and jth channels, assigning the jth channel to the ith horn and the ith channel to the jth horn, obtaining in place of Eq. (8a)

$$\left(\frac{P_i}{\sigma_j^2} + \frac{P_j}{\sigma_i^2} \right) = \frac{P_i}{(\sigma_i^2 + \Delta_{ij})} + \frac{(P_i - \delta_{ij})}{\sigma_i^2} \qquad (8b)$$

This pairwise exchange of a lower index channel with a higher index channel shall be termed an "unsorting exchange," for reasons that will soon become apparent. A comparison of the initial and exchanged pairs shows that the sum of the SNRs in Eq. (8b) is less than that in Eq. (8a) by an amount

$$l(i,j;i,j) = \frac{\Delta_{ij}\delta_{ij}}{\sigma_i^2(\sigma_i^2 + \Delta_{ij})} \geq 0 \qquad (9)$$

Because a pairwise exchange does not affect the rest of the channel assignments, it follows that the total SNR also decreases by exactly this amount as a result of the exchange.

Next, allow an arbitrary channel assignment (that is, not the "standard assignment" defined above), and perform an unsorting exchange on this configuration. This again leads to a decrease in the sum, now by an amount

$$l(i,j;m,n) = \frac{\Delta_{mn}\delta_{ij}}{\sigma_m^2(\sigma_m^2 + \Delta_{nm})} \geq 0 \qquad (10)$$

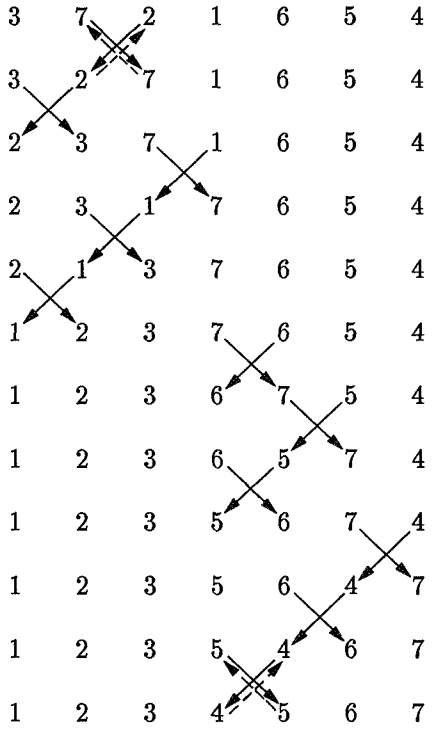
where m and n, $m < n$, are the indices of the exchanged channels (as before, i and j are the indices of the horns, in order of decreasing signal power). Thus, regardless of the state of the system, an unsorting exchange always leads to a decrease in the achievable combined SNR. If it can be shown that any channel assignment can be reached from the standard assignment by performing only unsorting exchanges, then it will have been proven that the standard assignment achieves the highest possible combined SNR for the given set of signal powers and noise variances.

This property can be shown in an interesting way, using the following (rather inefficient) sorting concept. The algorithm sorts a sequence of numbers by sequentially comparing nearest neighbors, starting at the left end of the sequence, and performing a pairwise "sorting" exchange if and only if the value of a given element exceeds that of its neighbor to the right (thus, for this algorithm, the indices i, j, and m, n are always consecutive integers). After each exchange, the algorithm restarts at the beginning of the sequence and continues to compare each element to its rightmost neighbor until no further exchanges are possible. At this point, the sequence is sorted with the smallest element appearing at the left. Retracing the exchange path

from the sorted sequence to its initial state shows that the "unsorting" operation requires only unsorting exchanges. The following example illustrates the procedure. Consider a 7-horn array with standard ordering, and let the channel assignment be denoted by some arbitrary index sequence, say

3 7 2 1 6 5 4

(that is, the channel with the third lowest noise variance assigned to horn number 1, with the greatest noise to horn number 2, and so on). Applying the above sorting algorithm to this sequence yields



where the downward pointing arrows indicate the pairwise "sorting" exchanges performed by the algorithm to arrive at the standard channel assignment. When these exchanges are applied in reverse, as indicated by the upward pointing arrows, it is clear that only unsorting exchanges were used to unsort the standard assignment (hence the terminology). Since the above algorithm sorts any sequence, it follows that one can arrive at any channel assignment by performing only unsorting exchanges on the receiver channels. Since each such exchange decreases the resulting combined SNR, it follows that the standard channel assignment achieves the highest combined SNR possible with any set of signal power and noise variances distributions.

B. Channel Assignment to Maximize ρ_{ML}

The standard ordering does not always maximize ρ_{ML}. This is demonstrated by the following example. Suppose that $\eta = 1000$, $\delta = 60$ deg, $K = 7$, and $L = 4$, so that the smallest number of samples is observed for which Eq. (1) is valid. Further suppose that the power levels in the first, second, ..., seventh horns are

10 8 7 6 6 5 5

while the corresponding noise variance distribution is

0.5 1 1 1 1 1 1

which implies a standard channel assignment. Direct substitution into Eq. (1b) and Eq. (2) yields

$$\rho = 28.5000 \qquad \rho_{ML} = 3.7635$$

Exchanging channel 1 with channel 2 yields the noise variance distribution

1 0.5 1 1 1 1 1

Recomputing Eq. (1b) and Eq. (2), one obtains

$$\rho = 27.5000 \qquad \rho_{ML} = 4.0480$$

which shows that ρ_{ML} increased while ρ decreased as a result of the exchange. Observe that this is not a preferred region of operation, since the combining losses are unacceptably high for such small L. However, the standard ordering does maximize ρ_{ML} in the limit as the number of observed samples grows without bound.

IV. Some Special Cases of Interest

Next, consider situations where the signal or noise distributions, or both, obey simple rules due to symmetries in the underlying model, namely:

(1) the noise variance is identical in $(K-1)$ of K channels (constrained channel noise model)

(2) the signal power is identical in $(K-1)$ of K channels (constrained signal distribution model)

(3) both the channel noise and the signal distribution are constrained according to (1) and (2) in the same $(K-1)$ channels (jointly constrained model)

The above conditions might not be strictly satisfied in practice, but the simplicity of the model may nevertheless provide useful approximations to actual operating conditions.

It will be assumed that independent estimates are made of the required parameters in each channel, without taking advantage of the special relations that exist when the above models hold. This is equivalent to admitting that the above special conditions are not known to exist a priori. If these conditions were known to exist a priori, the estimation algorithms could be matched to the unusually favorable distributions, resulting in improved performance. However, rather than pursue that idea, this article focuses instead on the performance of the "unmatched" independent channel estimators described in [1] and [2].

A. Constrained Channel Noise Model

Consider a model where the noise variance in one of the channels is considerably less than in the surrounding channels, all of which have the same (higher) noise variance. This situation could arise, for example, when a low-noise cooled maser is employed for reception and is surrounded by a ring of higher noise high-electron-mobility transistor (HEMT) low-noise amplifier (LNA) channels for possible gain improvement. The outer channels are intended for use at extreme antenna elevation angles, where mechanical distortions tend to spread the signal power in the focal plane.

Ignoring minor noise variations among the outer channels, the difference in channel noise variances can be accounted for by specifying the ratio of "low" to "high" noise temperatures: call this ratio α and let $0 < \alpha < 1$. Thus, the central channel is modeled as having the lowest noise temperature, while the surrounding channels are modeled as having the same higher noise temperature, each a factor $(1/\alpha)$ times that of the central channel. The two-sided spectral level of the thermal noise in the kth channel can be obtained from the corresponding noise temperature T_k as

$$\frac{N_{ok}}{2} = \frac{\kappa T_k}{2} \tag{11a}$$

so that for τ-sec sample averaging, the sample variance becomes

$$\sigma_k^2 = \frac{N_{ok}}{2\tau} \tag{11b}$$

(here κ is Boltzmann's constant). This model is sufficiently accurate to provide insight into the expected behavior of the combining system.

Although the central channel may have significantly less noise than the outer channels, it does not follow that only the central channel should be observed: there could be situations where the outer channels contain significant signal components that, when properly combined, could improve the overall SNR of the system. This situation may be due to distortions of the main reflector induced by gravity, wind gusts, defocusing of the subreflector, or even a simple pointing error. In each case, signal collected by the outer channels may improve the system SNR, as well as provide real-time pointing error information to the system.

Denote the sum of the channel SNRs by ρ, as before, and let ζ denote the sum of squares of channel SNRs. Thus, for the special case under consideration,

$$\rho = \sum_{k=1}^{K} P_k/2\sigma_k^2 = \frac{P_1}{2\sigma_1^2} + \frac{\alpha}{2\sigma_1^2}\sum_{k=2}^{K} P_k \tag{12a}$$

$$\zeta = \sum_{k=1}^{K} P_k^2/4\sigma_k^4 = \frac{P_1^2}{4\sigma_1^4} + \frac{\alpha^2}{4\sigma_1^4}\sum_{k=2}^{K} P_k^2 \tag{12b}$$

Letting Q_k denote the square of the kth signal power and Q_T denote their sum, Eqs. (12a) and (12b) can be rewritten as

$$\rho = \frac{\alpha P_T}{2\sigma_1^2} + (1-\alpha)\frac{P_1}{2\sigma_1^2} \tag{13a}$$

$$\zeta = \frac{\alpha^2 Q_T}{4\sigma_1^4} + (1-\alpha^2)\frac{Q_1}{4\sigma_1^4} \tag{13b}$$

which shows that ρ is a weighted average of P_1 and P_T. Thus, the effective SNR ranges from $P_1/2\sigma_1^2$ as α approaches zero (i.e., as the outer channels become infinitely noisy), to $P_T/2\sigma_1^2$ as α approaches 1.

B. Constrained Signal Distribution Model

Next, consider the signal distribution model. Let β be the fraction of the total received signal power intercepted by the central horn, $0 < \beta < 1$, and let the remaining signal power be distributed uniformly among the outer $(K-1)$ horns. Thus, if the total signal power intercepted by the array is P_T watts, and if the central horn is designated horn number 1 while the outer horns are labeled number 2–number K, the power in the central horn P_1 is

$$P_1 = \beta P_T \tag{14a}$$

while the power in any of the outer horns is

$$P_k = \frac{(1-\beta)}{(K-1)} P_T \tag{14b}$$

This model assumes a symmetrical signal distribution, which is generally valid for a focusing error, and may often be used to approximate the effects of gravity-induced mechanical distortions as well.

For this type of signal distribution (with arbitrary noise in each channel), one obtains

$$\rho = \frac{\beta P_T}{2\sigma_1^2} + \frac{(1-\beta)P_T}{2(K-1)} \sum_{k=2}^{K} \sigma_k^{-2} \tag{15a}$$

$$\zeta = \left(\frac{\beta P_T}{2\sigma_1^2}\right)^2 + \left[\frac{(1-\beta)P_T}{2(K-1)}\right]^2 \sum_{k=2}^{K} \sigma_k^{-4} \tag{15b}$$

The first term in Eq. (15a) is simply the SNR of the central channel. The combined SNR is augmented by an equivalent second channel with total signal power $\frac{(1-\beta)P_T}{K-1}$ and effective noise variance $\left(\sum_{k=2}^{K} \sigma_k^{-2}/2\right)^{-1}$

C. Jointly Constrained Model

When the above constraints on the noise variance and the signal distribution are simultaneously satisfied over the same channels, the expressions for ρ and ζ can be put into a particularly simple form:

$$\rho = \frac{P_T}{2\sigma_1^2}\left[\beta + \alpha(1-\beta)\right] \tag{16a}$$

$$\zeta = \left(\frac{P_T}{2\sigma_1^2}\right)^2 \left[\beta^2 + \frac{\alpha^2(1-\beta)^2}{(K-1)}\right] \tag{16b}$$

For the jointly constrained problem, there is only one nontrivial switching operation. It is instructive to examine the behavior of ρ and ρ_{ML} for standard and switched channel assignments in this case: let the subscript sw denote switched channel assignments. The behavior of these quantities, as well as their unswitched counterparts, is shown in Fig. 3, for the signal power distribution

100 10 10 10 10 10 10

and noise variance distribution

0.5 1 1 1 1 1 1

with $K = 7$, $\eta = 1000$, and $\delta = 80$ deg. Clearly, $P_T = 160$, $\beta = 10/16$, and $\alpha = 0.5$. Both ρ_{ML} and ${}_{sw}\rho_{ML}$ approach their limiting values for large L, but drop off rapidly below some critical number which, in this case, appears to be between 100 and 1000 samples. For even lower values of L, note that ${}_{sw}\rho_{ML} > \rho_{ML}$, indicating that the switching operation actually improves the combined SNR in this region. This behavior is attributed to the fact that in this region the switching operation generates so large a decrease in ζ that the value of the denominator in Eq. (2) decreases more than ρ, resulting in a net increase in ρ_{ML}. Although this unusual behavior is not expected to occur during normal operating conditions, it does point out a need to verify the optimality of the standard assignment in practice. The combining loss γ as a function of L corresponding to the standard assignment in the above example is displayed in Fig. 4. For $L > 1000$, the combining loss rapidly becomes negligibly small, reaching a value of less than 0.1 dB when about 3500 samples are observed.

V. Array Gain

Finally, consider the potential gain in SNR that could be achieved over a single feed horn by using an array of feeds in the focal plane. Using the notation of Eqs. (1) and (2), the array gain G_A is defined as the ratio of the combined SNR ρ_{ML} divided by the SNR in channel number 1:

$$G_A \triangleq \rho_{ML} / \left(P_1/2\sigma_1^2\right) = \rho/\gamma \left(P_1/2\sigma_1^2\right) \tag{17}$$

Since the combining loss γ is never less than 1, the array gain can be bounded with the ratio

$$G_A^u \triangleq \rho / \left(P_1/2\sigma_1^2\right) \tag{18}$$

which is recognized as the theoretical maximum of the array gain and achieved only with perfect weight estimates. The array gain and its upper bound may be expressed in decibels as

$$G_A(\text{dB}) = \rho(\text{dB}) - \left(\frac{P_1}{2\sigma_1^2}\right)(\text{dB}) - \gamma(\text{dB})$$

$$= G_A^u(\text{dB}) - \gamma(\text{dB}) \tag{19}$$

This expression shows that if the gain bound and the combining loss are known, the actual array gain can always be determined. To illustrate this point, the array gain, its upper bound, and the associated combining loss are shown in Fig. 5, for the example treated in Figs. 3 and 4. Although this example shows the behavior of G_A, G_A^u, and γ with increasing L, it corresponds to but a single point in (α,β) space. Greater appreciation for the benefits of array-feed combining may be obtained by examining the array gain as a function of the fractional signal power β for various noise temperature ratios α. This behavior is illustrated in Fig. 6. Since for large L the array gain approaches the gain bound G_A^u, only the gain bound is examined here, keeping in mind that the actual gain can always be obtained from these bounds by computing the combining loss and applying it to Eq. (19).

It is immediately apparent from Fig. 6 that G_A^u increases with decreasing β, which simply means that the importance of the outer ring increases as more signal power is diverted to it. It is also clear that for any β, the array gain increases with increasing α; that is, as the relative noise temperature of the outer ring decreases. This behavior is consistent with intuition, reaffirming that array feed combining is most effective when the signal power is spread over an area that is large as compared with the effective area of a single feed, and when similar quality channels are used to recover the signal. However, note also that considerable improvements are possible even in intermediate cases, for example, when $\alpha = \beta = 1/2$. This assignment refers to a situation where half the received signal power is diverted out of the central feed into the outer ring, while the noise variance in the central channel is half that of the outer channels (approximating the model for a hybrid maser/HEMT system operating with a distorted DSN antenna). The resulting array gain is seen to be close to 2 dB, which clearly justifies the use of an array combining system.

VI. Conclusions

The generalized results presented in [1,2] have been extended by allowing an additional maximization over the signal power and noise variance distributions in an array-feed combining system. It was shown that for any given signal power and noise variance distributions, a unique channel assignment exists that achieves the greatest combined SNR in the limit of vanishingly small combining losses, and that this limit can be approached in practice by observing a sufficient number of samples. Some special signal and noise distributions were examined, which approximated the signal power distribution characteristic of distorted antennas, observed by a low-noise channel augmented by a ring of higher noise receivers. It was shown that considerable gain improvements are possible by using such a "hybrid" arrangement. For example, even if half of the received signal power (3 dB) is diverted out of the low-noise channel into the outer ring, as much as 2 dB of SNR may be recovered by using receiver channels with twice the noise temperature of the low-noise receiver. This approach may prove to be a cost-effective way to improve the performance of low-noise (hence, high-cost) receivers degraded by distorted antennas operating at Ka-band (32 GHz) or higher carrier frequencies.

References

[1] V. A. Vilnrotter and E. R. Rodemich, "A Real-Time Signal Combining System for Ka-Band Feed Arrays Using Maximum-Likelihood Weight Estimates," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 81–99, February 15, 1990.

[2] V. A. Vilnrotter, E. R. Rodemich, and S. J. Dolinar, Jr., "Real-Time Combining of Residual Carrier Array Signals Using ML Weight Estimates," to appear in *IEEE Trans. on Comm.*, December 1991.

[3] P. W. Cramer, "Initial Studies of Array Feeds for the 70-Meter Antenna at 32 GHz," *TDA Progress Report 42-104*, vol. October–December 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 50–67, February 15, 1991.

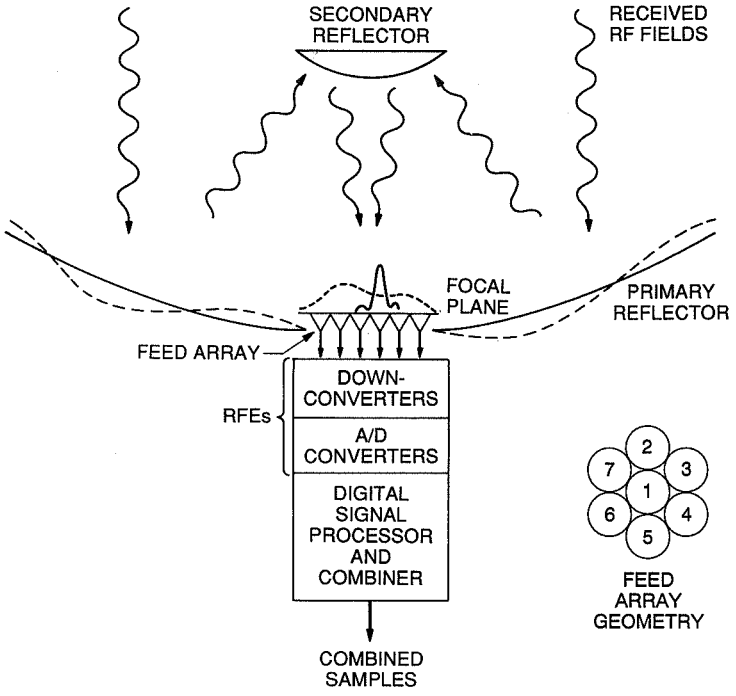


Fig. 1. Real-time antenna compensation system conceptual design.

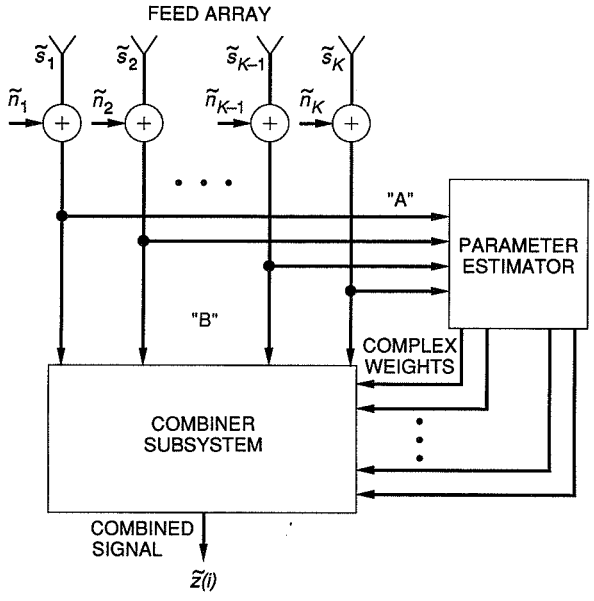


Fig. 2. Signal-combining system (baseband model) block diagram.

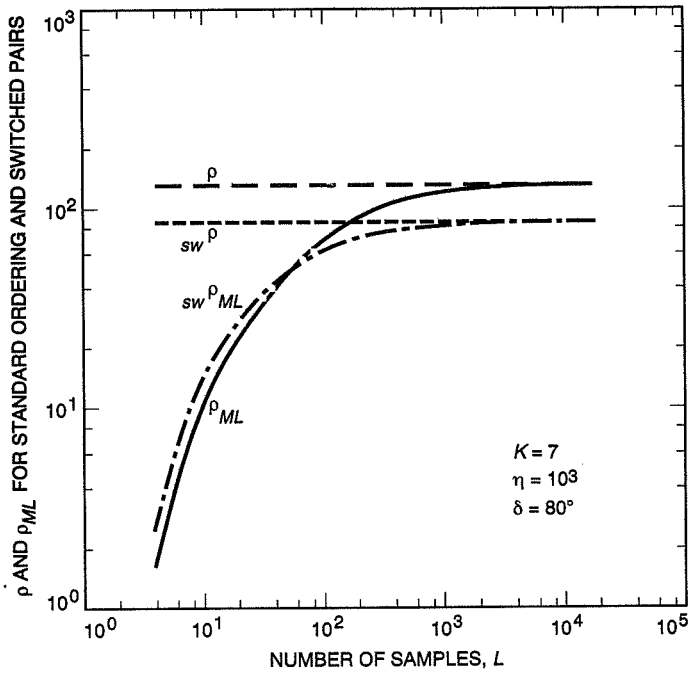


Fig. 3. Signal-to-noise ratios for standard and switched assignments versus *L*.

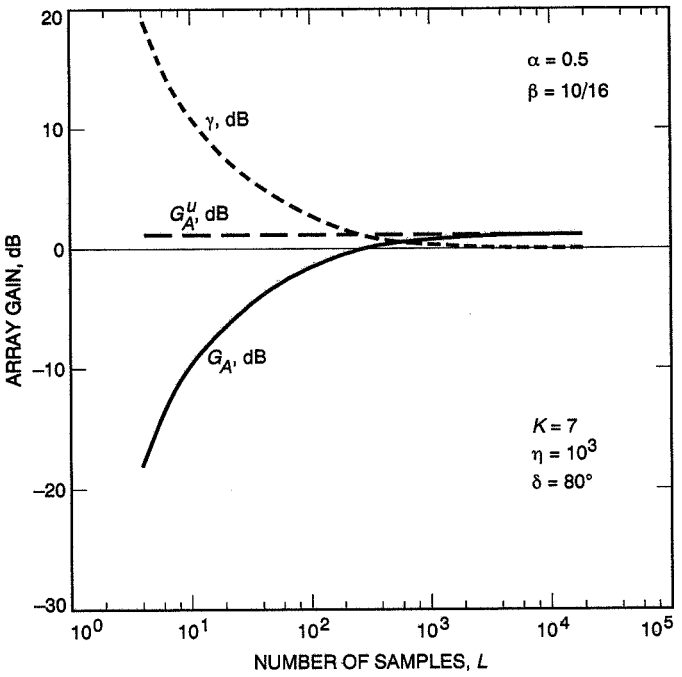


Fig. 5. Array gain and bound as a function of *L*.

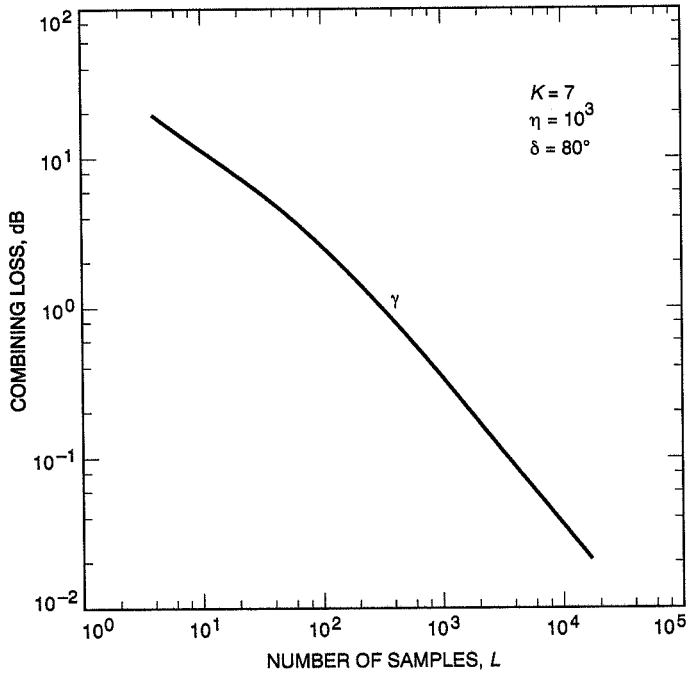


Fig. 4. Combining loss versus *L*.

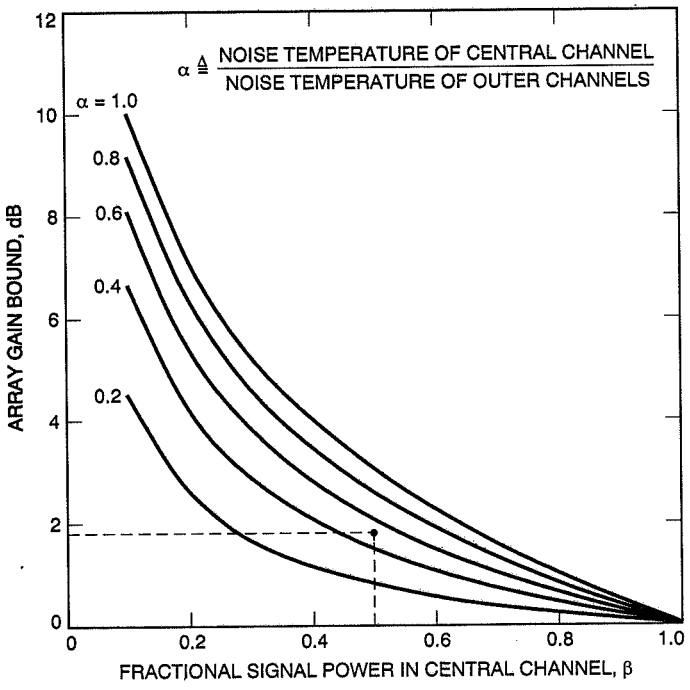


Fig. 6. Array gain bound as a function of α and β.

Coding Gains and Error Rates From the Big Viterbi Decoder

I. M. Onyszchuk
Communications Systems Research Section

A prototype hardware Big Viterbi Decoder (BVD) has been completed for an experiment with the Galileo spacecraft. This decoder resulted from six years of research and development by many members of the Communications Systems Research Section. Searches for new convolutional codes, studies of Viterbi decoder hardware designs and architectures, mathematical formulations and decompositions of the deBruijn graph into identical and hierarchical subgraphs, and VLSI chip design are just a few examples of tasks completed for this project. In this article, BVD bit error rates, measured from hardware and software simulations, are plotted as a function of bit signal-to-noise ratio E_b/N_0 on the additive white Gaussian noise channel. Using the constraint length 15, rate 1/4, experimental convolutional code for the Galileo mission, the BVD gains 1.5 dB over the NASA standard (7,1/2) Maximum-Likelihood Convolutional Decoder (MCD) at a bit error rate (BER) of 0.005. At this BER, the same gain results when the (255,223) NASA standard Reed-Solomon decoder is used, which yields a word error rate of 2.1×10^{-8} and a BER of 1.4×10^{-9}. The (15,1/6) code to be used by the CRAF/Cassini missions yields 1.7 dB of coding gain. These gains are measured with respect to symbols input to the BVD and increase with decreasing BER. Also, 8-bit input symbol quantization makes the BVD resistant to demodulated signal-level variations which may cause 0.1-dB or more loss for the MCD. Since the new rate 1/4 or rate 1/6 codes require higher bandwidth than the NASA (7,1/2) code, these gains are offset by about 0.1 dB of expected additional receiver losses. Coding gains of several decibels are possible by compressing all spacecraft data.

I. Introduction

The Big Viterbi Decoder (BVD) is a programmable, error-correcting machine for convolutional codes with constraint length from 2 to 15 and rate $1/2, 1/3, 1/4, 1/5$, or $1/6$ [7]. Recently, a prototype hardware unit was completed with an $8.\bar{3}$ MHz system clock and tested at 400,000 or more decoded bits/second. Implementation versions of this decoder will be installed at DSN stations to support the Galileo, CRAF/Cassini, and future missions. Searches

for new convolutional codes having constraint length 13 to 15 demonstrated that 1.5 to 2.0 dB of coding gain is possible [3,4]. A novel hardware design was completed for a constraint length 15, parallel Viterbi decoder [7]. The arithmetic core contains 8192 simple processing units, interconnected according to a hierarchical decomposition of the deBruijn graph into identical subblocks [8].

The (15,1/4) code described in [4] achieves a 1.5-dB coding gain over the (7,1/2) Maximum-Likelihood Convolutional Decoder (MCD), at a decoded bit error rate (BER) of 0.005, which corresponds to an outer NASA (255,223) Reed-Solomon decoder word error rate of 2.1×10^{-8} and BER of 1.4×10^{-9}. The gain is 1.45 dB at a Reed-Solomon BER of 10^{-6}. The (15,1/6) code described in [3] will be used by the CRAF/Cassini missions for a 1.7-dB gain over the MCD. These gains compare well with the 1.9-dB gain resulting from the upgrade of all three 64-m DSN antennas to 70 m, with surface shaping and improved maser. Also, the coding gains apply for all antennas utilized. On the other hand, compressing all spacecraft data 3:1, perhaps noiselessly, would yield a 4.77-dB gain, less 0.25 dB for increased Reed-Solomon error correction to counteract error propagation.

This article contains bit error rates for the BVD obtained by software simulations of the decoder hardware which were verified by running the BVD hardware on pseudo-random data with added Gaussian noise. Frame error rates for the NASA (255,223) Reed-Solomon decoder with inner BVD are given in another article [9]. All Reed-Solomon error rates given in this article are calculated assuming that symbols from many (i.e., 5 or 8) different codewords are interleaved prior to convolutional encoding. This ensures that symbol errors, which are caused by channel noise, within any particular codeword are statistically independent. For all simulations, decoder "node synchronization" was assumed, meaning that groups of n incoming symbols (where $1/n$ is the code rate) are aligned with the code trellis. This is realistic for the operating E_b/N_0 range and error rates of interest. Algorithms are now being developed to acquire and track node synchronization (i.e., to lock the BVD onto the incoming symbol stream) at BERs up to 0.1. The simulations are described in Section II of this article. Coding gains with respect to the (7,1/2) MCD are discussed in Section III. In Section IV these results are shown to be accurate to within 0.03 dB.

II. Decoder Simulations

Bit error rates as a function of information-bit signal-to-noise ratio E_b/N_0 were obtained from software simulations of the actual BVD hardware. The software program is similar to those used to actually debug and test the BVD. An encoded pseudo-random information sequence with added, precisely controlled, wideband Gaussian noise was decoded. A linear feedback shift register with period $2^{31}-1$ was used to generate 5 million to 60 million data bits for these tests. Each bit yielded n channel symbols output by a convolutional encoder, where $1/n$ is the code rate. Symbols 0 and 1 were mapped to demodulated signal levels of +0.84 V and −0.84 V, respectively. Analog, wideband Gaussian noise was then added to these channel symbols. Exactly as in the baseband assembly (BBA) or symbol synchronizer assembly (SSA), demodulated channel symbols (which include noise) were truncated to ±5 V and then quantized to 8 bits by using a step size of 0.0390625 V. The above procedure of adding noise to constant channel symbols corresponds to optimal SSA/BBA operation. In the BVD hardware simulations, noise samples were quantized before addition to channel symbols, a reversal of operations which makes no significant impact on the Viterbi decoder.

The noise source is based upon a standard multiplicative linear congruential generator for random uniform deviates [6]:

$$x_{n+1} = 16{,}807 x_n \pmod{2^{31}-1} \qquad n \geq 0$$

$$x_0 = 1099 \qquad \text{for initial seed}$$

$$u_{n+1} = -1 + 2x_{n+1}/(2^{31}-1)$$

The sequence of u_n's is uniformly distributed on the open interval $(-1, +1)$. If u_n and u_{n+1}, n odd, both lie within the unit circle, then a Box-Muller transformation yields two independent, normal (Gaussian) random deviates [2], which are multiplied by the desired standard deviation of the additive white Gaussian noise to produce two noise samples for addition with encoded symbols. Although the resulting noise spectrum has nearly flat magnitude, there are significant variations, perhaps because samples output by the linear congruential generator are weakly correlated [5]. Using a 4096-entry shuffling array on samples output by the linear congruential generator flattened the noise spectrum and increased BER, but only in the third significant digit—probably because encoded channel symbols were random and because Viterbi decoders are robust with respect to input signal plus noise. Nonetheless, careful selection of a wideband Gaussian noise source ensured accurate BER measurements. The effect on Viterbi decoding of filtering in the receiver may be analyzed in the future.

III. BVD Coding Gain Over the (7,1/2) MCD

Software BVD bit error rates for several convolutional codes are in Fig. 1, where straight lines connect raw data points. BVD hardware runs, using 5 million bits of repeating pseudo-random data with period 65,535, are also in Fig. 1, and they match the software data to within 0.05 dB. For the hardware tests, E_b/N_0 values were calculated from the 8-bit quantized noise samples added to the encoded symbol stream.

The nominal operating point of a Viterbi decoder in the DSN receiving stations is BER equal to 0.005. At this error rate, coding gains of 1.5 and 1.7 dB are achieved by the BVD with a (15,1/4) or (15,1/6) code, respectively, over the (7,1/2) MCD. These same gains occur when an outer (255,223) Reed-Solomon code is used [9]. The BVD uses all 8 bits of input symbols, instead of only the 3-bit quantization used by the MCD. This feature results in a 0.16-dB gain for the (7,1/2) BVD over the (7,1/2) MCD. It also makes the BVD resistant to variations in the mean absolute symbol level, which is controlled by the BBA/SSA. Such variations may cause a loss of 0.1 dB or more for the MCD. Notice in Fig. 1 that the BVD coding gains increase with decreasing BER. For example, at BER $= 10^{-3}$, the BVD gains 1.7 dB by using the (15,1/4) code and 1.9 dB by using the (15,1/6) code. The corresponding (255,223) Reed-Solomon decoder BER is 10^{-19}.

An (11,1/6) decoder is implemented with one processor board and a 552-wire connector, instead of 16 boards connected by a 28-layer, 4416-wire printed circuit backplane in the full BVD for (15,1/6) codes. The BER curves for an (11,1/6) code show that there is less than 0.09 dB of gain possible for each increase by 1 in encoder constraint length (which doubles the amount of decoder hardware) from 11 to 15 or higher, at which point the gain saturates rapidly. However, all hardware complexity in the BVD may be justified by the economics of upgrading antennas.

The formula[1]

$$\text{BER} \approx \exp(9.807 - 14.064 E_b/N_0)$$

was used for BVD bit error rates for the (15,1/6) code from $E_b/N_0 = -0.1$ to 0.5 dB. (But the formula requires E_b/N_0 to be an arithmetic ratio, not in logarithmic decibel units.) This approximation yields a curve right on top of the one for hardware (15,1/6) BERs in Fig. 1. The formula

$$\text{BER} \approx \exp(10.11 - 13.68 E_b/N_0)$$

was obtained by curve-fitting the BVD (15,1/4) code software simulation data in Fig. 1. Similarly, the formula

$$\text{BER} \approx \exp(4.55 - 6.23 E_b/N_0)$$

was obtained from the BVD (7,1/2) code software simulation data. The last two formulas both match the corresponding software data curves in Fig. 1. All approximations above require that E_b/N_0 be expressed as an arithmetic ratio instead of in logarithmic decibel units.

IV. Accuracy of the Results

Errors from a Viterbi decoder occur in bursts of output bits, during which each decoded bit is correct with a probability of nearly 1/2. The error statistics of the (7,1/2) and (7,1/3) MCD have been analyzed in great detail [1]. That report contains an estimate of the uncertainty in BER measured from simulations. Let σ_{BER} denote the variance in BER caused by different possible noise sequences and by a limited number of error bursts during a simulation. Then from [1],

$$\frac{\sigma_{\text{BER}}}{\text{BER}} \approx \frac{\alpha}{\sqrt{nbursts}}$$

where *nbursts* is the number of error bursts and α is a small constant that depends upon E_b/N_0 and the code. Simulations for the (7,1/2), (15,1/4), and (15,1/6) codes verified that the factor α is a constant less than 1.4, and decreases to 1.0 at high E_b/N_0. Hence, $\alpha = 1.4$ will be used herein and the ratio $\alpha/\sqrt{nbursts}$ will be called the relative BER error. For a fixed number of decoded bits, the largest relative BER error will occur at the highest E_b/N_0, when there will be fewest error bursts.

The formulas given in Section III may be used to convert relative BER error into uncertainty in E_b/N_0 for a given, fixed BER. For the BVD (7,1/2) software simulations, the relative BER error was a maximum of 0.022, which occurred at $E_b/N_0 = 3.0$ dB. By using the approximation formula, this error corresponds to uncertainty in E_b/N_0 of 0.015 dB. At $E_b/N_0 = 1.0$ dB, the relative BER error was 0.01115, which corresponds to an uncertainty in E_b/N_0 of 0.013 dB. For the (15,1/6) and (15,1/4) code software simulations, the E_b/N_0 uncertainty was typically 0.03 dB and at most 0.05 dB.

[1] S. Dolinar, "Empirical Formula for the Performance of the Recommended (15,1/6) Convolutional Code," Interoffice Memorandum 331-90.2-060 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 12, 1990.

Acknowledgments

The author thanks J. Statman and all BVD team members for their support during system integration, and for providing the means to obtain the hardware simulation data. The author is also grateful to G. Zimmerman for helpful suggestions during noise analysis.

References

[1] R. L. Miller, L. J. Deutsch, and S. A. Butman, *On the Error Statistics of Viterbi Decoding and the Performance of Concatenated Codes,* JPL Publication 81-9, Jet Propulsion Laboratory, Pasadena, California, September 1, 1981.

[2] D. E. Knuth, *Seminumerical Algorithms,* 2nd ed., vol. 2 of *The Art of Computer Programming,* Reading, Massachusetts: Addison-Wesley, pp. 117–118, 1981.

[3] J. H. Yuen and Q. D. Vo, "In Search of a 2-dB Coding Gain," *TDA Progress Report 42-83,* vol. July–September 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 26–33, November 15, 1985.

[4] S. Dolinar, "A New Code for Galileo," *TDA Progress Report 42-93,* vol. January–March 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 83–96, May 15, 1988.

[5] P. L'Ecuyer, "Efficient and Portable Combined Random Number Generators," *Communications of the ACM,* vol. 31, no. 6, pp. 742–749, June 1988.

[6] S. K. Park and K. W. Miller, "Random Number Generators: Good Ones Are Hard to Find," *Communications of the ACM,* vol. 31, no. 10, pp. 1192–1201, October 1988.

[7] J. Statman, G. Zimmerman, F. Pollara, and O. Collins, "A Long Constraint Length VLSI Viterbi Decoder for the DSN," *TDA Progress Report 42-95,* vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 134–142, November 15, 1988.

[8] O. Collins, F. Pollara, S. Dolinar, and J. Statman, "Wiring Viterbi Decoders (Splitting deBruijn Graphs)," *TDA Progress Report 42-96,* vol. October–December 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 93–103, February 15, 1989.

[9] L. Swanson and K.-M. Cheung, "Frame Error Rate of the NASA Concatenated Coding System," *TDA Progress Report 42-106,* vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 183–187, August 15, 1991.

C-3

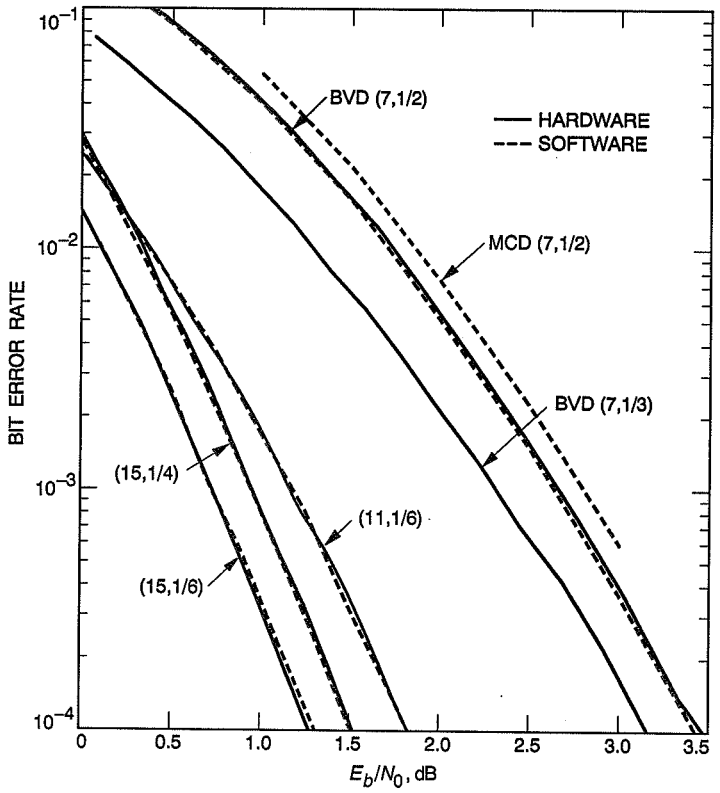


Fig. 1. Viterbi decoder bit error rates for several codes.

Testing Interconnected VLSI Circuits in the Big Viterbi Decoder

I. M. Onyszchuk
Communications Systems Research Section

The Big Viterbi Decoder (BVD) is a powerful new error-correcting hardware device for the Deep Space Network (DSN), in support of the Galileo and CRAF/Cassini missions. Recently, a prototype was completed and run successfully at 400,000 or more decoded bits per second. This prototype is a complex digital system whose core arithmetic unit consists of 256 identical very large scale integration (VLSI) gate-array chips, 16 on each of 16 identical boards which are connected through a 28-layer, printed-circuit backplane using 4416 wires. Special techniques were developed for debugging, testing, and locating faults inside individual chips, on boards, and within the entire decoder. The methods are based upon hierarchical structure in the decoder, and require that chips or boards be wired themselves as Viterbi decoders. The basic procedure consists of sending a small set of known, very noisy channel symbols though a decoder, and matching observables against values computed by a software simulation. Also, tests were devised for finding open and short-circuited wires which connect VLSI chips on the boards and through the backplane.

A new "single-board" BVD is being constructed now for implementation in DSN stations. The design is based upon a new VLSI chip which contains 65,536 bits of static RAM and circuitry now located on boards and in 4 gate arrays. Sixty-four of these new chips will be mounted on a single board and interconnected with 5120 wires. Experience gained from the prototype actually led to special circuitry designed inside the chips for simpler and better checking of these wires. Furthermore, each chip also operates independently as a complete, constraint length 9, rate 1/6—(9,1/6) Viterbi decoder, for testing within the single-board BVD. Like the prototype BVD, the new decoder can also be programmed to decode convolutional codes having constraint length 2 to 15 and rate 1/2, 1/3, 1/4, 1/5, or 1/6.

I. Introduction

A Viterbi decoder is a probabilistic device which may be ultimately verified by measuring bit error rates over a range of channel noise levels [3]. However, such a test requires considerable work for a simple pass/fail result, and no information about the source of error is obtained when the decoder fails. Therefore, other techniques were required to test the Big Viterbi Decoder (BVD), and to pinpoint sources of error when the decoder stops working. It was discovered that running the decoder with a small number of known, very noisy channel symbols was extremely effective for testing and isolating faults within the BVD. Furthermore, the modular construction of the BVD enabled independent checking of individual chips and boards as smaller Viterbi decoders prior to their installation in the BVD.

The problems addressed in this article are testing, fault location inside, and debugging of the BVD. The problems also include checking of "metric connections"—not control signals—between very large scale integration (VLSI) chips. A connection is defined as a chip output driver, the "metric" wire it drives to another chip, and the corresponding input driver on the other chip. There are 4416 such connections in the prototype BVD. Unique software programs were required to verify connections on boards and in backplanes. These procedures were used to debug the BVD prototype and are now part of the system software.

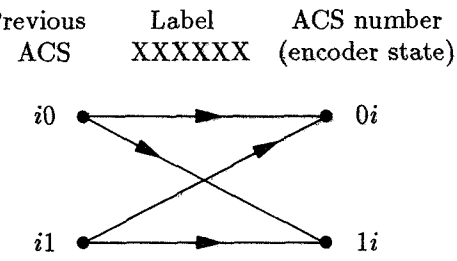
The basic idea is to connect one or more chips in order to implement the arithmetic core of a parallel Viterbi decoder. The hierarchical decomposition of the (15,1/6) BVD into identical sub-blocks [2] makes this approach possible. For example, one gate array may be wired as the arithmetic part of a (7,1/6) decoder. Similarly, each of the 16 boards in the prototype BVD, which contain 16 gate arrays, may be tested independently as the core of a (11,1/6) decoder. Thus, chips and boards may themselves be individually and effectively tested as smaller Viterbi decoders before they are integrated into the BVD.

This article is organized as follows. Basic arithmetic processors called "butterflies," and their interconnection in the BVD, are described in Section II. Procedures for testing the VLSI chips are explained in Section III. Techniques are described in Section IV for locating short-circuited and open-circuited metric wires on boards and in the backplane. The single-board BVD is described in Section V. Based upon experience with the 16-board prototype, a plan is outlined for independently testing new VLSI chips and connections in the single-board BVD.

II. Butterflies

The fundamental arithmetic unit in the BVD, a "butterfly," is described in this section from a purely operational viewpoint. The following explanation should be sufficient for understanding the remainder of this article. The basic theory underlying the use of butterflies in Viterbi decoders appears in previous articles [1,2].

The "butterfly" shown below is the basic arithmetic unit in the BVD [1,2]:



$B = 2^{K-2}$ butterflies may be connected to form the arithmetic core of a parallel, rate $1/n$, constraint-length K, Viterbi decoder [2]. For example, 32 butterflies are wired in Fig. 3 to form the arithmetic part of a $K = 7$ decoder. Each butterfly in the decoder is identified by a number, i, in the range 0 to B–1 inclusive. Butterfly i has two processors, called Add-Compare-Select (ACS) units for reasons that will become evident when the ACS function is described later. The ACS units are marked $0i$ and $1i$ for the encoder states to which they correspond [1]. An ACS unit in the decoder and the corresponding encoder state will be referred to interchangeably. After the hardware comprising a butterfly is reset, but before the butterfly starts operating, a 6-bit "label" is loaded into a register inside the butterfly. This label is computed using the butterfly's number (i) and 6 encoder generator polynomials [1]. If the code has less than 6 generator polynomials, because the code rate is not 1/6, then zero polynomials are added.

A decoder cycle will denote the basic unit of processing time for a Viterbi decoder. During each decoder cycle, butterfly i receives two 16-bit numbers, called state metrics, bit-serially (least significant bit first): one from the output of ACS $i0$ in butterfly $2i \pmod B$ and the other from ACS $i1$ in butterfly $2i + 1 \pmod B$. The state metrics, stored inside ACS units, are all set to a common value such as 0 when the butterflies are reset. During each decoder cycle, all butterflies also receive a common set of six 8-bit quantized, demodulated channel symbols. These are used with the label inside each butterfly to compute a number (branch metric) for each of the 4 branches. Now for each of the two incoming branches, each ACS unit adds the branch

metrics to the corresponding state metric. Each ACS unit keeps only the smaller of the two resulting numbers, for transmission as a state metric input to a butterfly during the next decoder cycle, and writes into a random-access memory (RAM) the least significant bit of the ACS/state whose metric led to the smallest number. The bits written to RAM are called traceback bits because during later decoder cycles, a process traces backwards through the RAM to extract the decoded bits to be output by the Viterbi decoder. During every decoder cycle, each one of the $2B$ ACS units writes a 0 or 1 into a particular location in traceback memory (Fig. 1). The location is indexed by two numbers: a state number which is the ACS unit number, and a column number which is the same for all ACS units and depends only on the decoder cycle. After each decoder cycle, one column of traceback RAM is written. Starting from reset, columns 0 to 509 are written sequentially, but, thereafter, a column is overwritten immediately after a traceback process has used it.

In this article, "testing" a chip, board, or decoder means verifying that no manufacturing or assembly defects exist. "Fault isolation" means precisely locating sources of error when the hardware stops operating correctly. "Debugging" means finding and fixing errors in logic design and timing. Techniques based upon software simulations using known channel symbols, corresponding to an information bit signal-to-noise ratio E_b/N_0 of 0 dB, were instrumental during all phases of BVD testing, debugging, and fault isolation.

For purposes of testing a butterfly, the controllables are a 6-bit label which is loaded once after reset, and then groups of six 8-bit symbols. The only observables are traceback bits written from the butterflies into RAM. However, during the debugging of chips and boards, a logic analyzer was required to observe state metrics.

III. Testing VLSI Chips

Before encapsulation inside packages with pins, VLSI chips are tested by applying a sequence of binary patterns called "test vectors" to input pads during successive clock cycles. The test vectors are designed to exercise all circuitry inside the chip and expose any faults created during fabrication. The resulting vectors of bits at chip output pads are compared with expected, correct vectors obtained from a software simulation, like one used to verify logic and timing within the chip. Two different VLSI chips were designed for the BVD: a semi-custom, military-specified chip containing 16 identical butterflies, circuitry for processing control signals, and circuits for organizing traceback bits before they are written to memory; and a commercial gate array which implements two semi-custom chips, for a total of 32 butterflies.

The first VLSI testing for the BVD involved the semi-custom chips. Every flip-flop inside these chips, except those in the middle of serial registers, has a two-input multiplexor at the data input. All of these multiplexors are controlled by a common TEST signal. When TEST is active, each multiplexor selects the output of another flip-flop, instead of a combinational circuit output as it would during normal operating mode. Thus, all flip-flops inside the chip may be chained together in order to load or read bit patterns. In order to test a chip, the procedure described below is repeated once for each pattern in a very carefully designed set.

A pattern is loaded bit-serially into flip-flops while the TEST signal is active. The chip is then clocked for one or more cycles while TEST is inactive. Next, the contents of all flip-flops are read out as a serial bit stream while TEST is active, and compared with a known correct pattern in order to detect faults. The total number of clock cycles needed for each bit pattern is a few more than twice the number of flip-flops in the chip. However, the number of test vectors required to implement all bit patterns in the designed set can be very large, particularly if there is only one serial chain path which connects all flip-flops in the chip, instead of many parallel chains—for example, one for each butterfly. There is only one serial chain path in the semi-custom chips, and it joins all 1776 flip-flops in the chip. Furthermore, the chip requires two clock cycles per test vector. Hence, at least 7106 test vectors are required for each bit pattern. However, integrated circuit test equipment for both the semi-custom and gate array BVD chips could process at most 16,384 different test vectors. Furthermore, any testing of chips individually will not expose potential problems when several identical chips are wired together. When 512 semi-custom chips are installed in the BVD, fault isolation using bit patterns which are serially loaded and unloaded may take a long time and require complex software for analyzing output bits.

The semi-custom chips were tested by the manufacturer using a set of test vectors created by a software simulation of a single butterfly. The 4 metric inputs and outputs for each butterfly inside these chips are wired directly to pins. All 16 butterflies were tested in parallel, using the same label, noisy symbols, and input metrics. However, it was not certain, for reasons explained below, that the chips could be connected as the ACS part of a Viterbi decoder. Hence, the 32 butterflies in two semi-custom chips were interconnected as shown in Fig. 3.

A software simulation was run for the two interconnected chips. The program evolved from one used to gen-

erate test vectors for the gate arrays. Pseudo-random data were sent through a (7,1/6) convolutional encoder and the 0/1 bits output were mapped to ±21, which represent quantized, demodulated channel symbols when there is no channel noise. Then 8-bit quantized, zero-mean, white Gaussian noise was added to these symbols. The noise variance was set for a bit signal-to-noise ratio of 0 dB, an extremely difficult operating point for a (7,1/6) Viterbi decoder. Groups of 6 noisy symbols were processed by the software simulator, and test vectors were generated. The pair of semi-custom chips passed the test vectors (herein called the zero-dB check) only after the test vectors were modified as described below. Due to timing of signals at chip pins, every state metric became hex 4000 instead of 0000 after reset. This anomaly does not affect traceback bits, so it merely resulted in a simple change to the software simulation. A few other problems found were compensated for outside the chips. These small discrepancies could not be exposed by the original set of test vectors, but understanding the anomalies turned out to be crucial during later BVD testing.

The zero-dB check demonstrated that two semi-custom chips could work as the ACS part of a (7,1/6) Viterbi decoder. The results in [2] showed that 2^{K-6} properly interconnected chips would form a (K,1/6) decoder, for $K \geq 6$. These facts inspired confidence that 32 semi-custom chips on a board could be tested as the ACS part of an (11,1/6) decoder, and that 16 boards connected through a backplane would implement the ACS section of the (15,1/6) BVD.

A batch of gate-array chips was tested at 25 MHz by the manufacturer using test vectors generated by a simulation like that used for the zero-dB check, except that extremely noisy symbols (corresponding to E_b/N_0 near -8 dB) were used to better exercise butterfly logic. Twenty chips that passed tests were packaged as engineering prototypes. The 32 butterflies inside each of these gate arrays were then connected using a test fixture with wires connecting chip pins as the ACS part of a (7,1/6) Viterbi decoder. All 20 gate arrays passed the zero-dB check. As a result, 600 chips were ordered to build two prototype BVDs, each one with 256 gate arrays, 16 gate arrays on each of two spare boards, and 12 spare chips. The zero-dB check was also used to test individual boards, each with 16 gate arrays, when the boards were wired using a special edge connector as the ACS part of an (11,1/6) Viterbi decoder.

IV. Locating Bad Metric Wires (Connections)

During BVD debugging, procedures were needed to verify that VLSI chips already installed in the system were still working, and that all connections between chips were correct. Thus, a "traceback read" procedure was implemented, based upon the zero-dB check, to match a group of traceback bits written to the entire traceback memory, against values computed during a software simulation. Also, new procedures were devised for locating broken or missing metric wires (open circuits) and short-circuited metric wires in the BVD. These tests (1) examine traceback bits computed inside butterflies and stored in memory, (2) were required to debug backplanes, and thus (3) are now part of the BVD software. All of the above tests start by resetting the decoder and then loading a particular set of butterfly labels into the chips.

In the traceback read procedure, 510 sets of 6 symbols are processed so that every traceback memory location is written to and later read from. However, only the first wrong traceback bit can be reported, even though state metrics could be incorrect during decoder cycles prior to the one for which the first wrong bit is computed. Also, reading traceback memory cannot always be used to determine which of the two metrics input to the butterfly, for which a traceback bit error is reported, is incorrect. This problem could be reduced by repeating the procedure using many different sets of symbols, which also helps to detect intermittent faults. A traceback bit error could be caused by one of the chips driving metric wires, by connections between chips, or by the chip receiving the two metrics (Fig. 2). Without a tester for chips, it is impossible to isolate the fault, so the receiving and then driving chips are replaced in that order. If this does not solve the problem, the connection between chips is suspected.

The procedures described next require that the VLSI chips, memory chips, and interface circuitry on every board operate perfectly. This is verified in the prototype (16-board) BVD by testing each board independently as an (11,1/6) decoder.

A. Detecting Open-Circuited/Grounded Metric Wires

In the prototype BVD, or any subset of it wired as a (K,1/6) Viterbi decoder, if a metric wire is open-circuited or grounded, the metric value will remain zero if the wire floats low, which is usually the case because there is nothing to drive the wire. This zero metric will propagate, with branch metrics added, to all state metrics after $K-1$ sets of 6 symbols are processed. The zero metric will continue to corrupt every other state metric.

Detecting open-circuit metric wires is straightforward: the label 010101 is loaded into each butterfly, the decoder is reset, and six particular symbols are processed to force

every ACS unit to select the "horizontal" branch input. Then for each state, a "traceback" bit is read from column 0 of the memory and an error reported when the bit differs from the most significant bit of the state. In the case of an error, a diagonal branch was chosen so the state having a wrong traceback bit is reported, as well as the butterfly input 0 or 1 which caused the error. A similar procedure detects metric wires connected to power.[1]

B. Detecting Short-Circuited Metric Wires

When two or more metric wires are connected together, the resulting signal has an intermediate voltage (between 0 and 5 volts) during arithmetic clock cycles in which the metric values output by VLSI chips drive the line differently (recall that metrics are processed bit-serially). The resulting metric value may be lower, equal to, or higher than the correct value. However, when the wire is being driven high and low by two different metric output drivers, a logic low is usually interpreted by VLSI chip inputs. The following example demonstrates a possible resulting metric C when two state metrics, labelled A and B, are shorted together:

A : 0 1 0 0 0 0 0 1 1 1 0 0 0 1 0 1

B : 0 1 0 0 0 0 1 0 0 1 1 0 0 0 1 1

C : 0 1 0 0 0 0 0 0 1 1 0 0 0 1 0 1

While an open-circuit metric may be detected during the first set of traceback reads, which occurs from column 0 of memory, shorts are detected only after the second set of symbols has been processed because all metrics are the same constant after reset, so all column 0 traceback bits are correct if no wires are open and all circuitry works properly. Hence, no errors must be reported during an open-circuit metric test prior to running the short-circuit test described below.

The algorithm described below locates with very high probability all metric wires having multi-level signals, but it cannot determine which wires are shorted together. Consider the set of 64 groups of six symbols, formed by taking a given group of 6 noisy symbols and inserting all 64 possible 6-bit vectors in place of the 6 sign bits. For each group of 6 symbols in this set, the decoder is reset, 6 fixed noisy symbols are processed, and then the group is sent to the decoder. Every bit in column 1 of traceback memory is read to determine which metrics are corrupted by shorts (recall that column 0 is correct because the decoder passed the open-circuit test). When an incorrect traceback bit is read from memory, the butterfly input corresponding to this bit is assumed shorted because a corrupted metric usually has a lower value, so the decoder selects it instead of the correct path. After all 64 iterations are complete, one for each group of 6 symbols in the set of 64, the number of errors found at each of the two metric inputs is reported for every decoder state.

For all runs of this algorithm on the BVD hardware, a multi-level signal has always been found on the metric input for which the largest number of errors was reported at a given state. In fact, one of the two metric inputs to a state having wrong traceback bits nearly always had 0 errors reported. The 64 groups of symbols have the effect of creating all 64 possible sets of 4 branch metrics inside a butterfly for 6 given input symbols. These different branch metrics help expose erroneous state metrics—sent along wires between chips—by propagating errors to the traceback bits, which are the only observables. Of course there is no guarantee that the algorithm will expose every short, but it has always worked in practice. In fact, the algorithm located touching pins on the backplane and in extender cards, caused by misalignment when boards were installed. In order to increase the probability that a short-circuited metric wire is detected, the above algorithm could be repeated with one or more additional, completely different symbol sets and the errors accumulated before final reporting.

V. Testing the Single-Board BVD

For implementation in DSN stations, the 16 boards with 256 identical 32-butterfly gate arrays in the prototype BVD are being compacted into 64 identical new VLSI chips to be mounted on a single backplane (Fig. 4). Each new chip contains 128 butterflies, 65,536 bits of static RAM for traceback, and circuitry for traceback, decoding, and testing. These will be interconnected with 5120 wires on a single board. Key system design objectives include automatic checking of all metric wires between chips and testing chips individually, perhaps using on-chip self-test circuitry not yet designed, while the chips are installed in the decoder. Due to the new VLSI chips, the arithmetic core of the single-board BVD will be more reliable and about 4 times less expensive than the arithmetic core in the current BVD prototype. Furthermore, the maximum rate of the single-board BVD will be at least 1.5 million bits/sec, at a 20-MHz system clock.

[1] I. M. Onyszchuk, "Finding Open and Short-Circuited Metric Wires in the BVD," Interoffice Memorandum 331-91.2-014 (internal document), Jet Propulsion Laboratory, Pasadena, California, March 18, 1991.

Each of the new VLSI chips can operate as a complete (9,1/6) Viterbi decoder. Test circuitry has been designed to internally connect all 128 butterflies to form the ACS section. Every cell in the traceback memory can be checked individually. Thus, even while installed in the BVD, each chip may be tested separately. Of course, on-chip self-test circuitry, not currently part of the new chips, would also enable such testing; each chip could be commanded using a special signal to test its internal circuitry and memory, and then provide a pass/fail signal. Chip self-testing has the advantage of verifying correct operation using only 2 signals. Testing a chip using external symbols and signals requires perfectly working interface circuitry. However, none of the input or output drivers for metrics would be tested by either of these techniques. The method suggested below will expose a bad driver by using multiplexors inside ACS units in order to check metric wires between chips.

Reading traceback memory is a difficult and time-consuming operation in the current prototype BVD. Although the tests have been successful and instrumental in fault isolation, they are not foolproof and require manual interpretation. Therefore, a better method is required to test the decoder and to locate faults. Also, 160 of the 196 signal pins on the new VLSI chips are used for metric wires. Thus, connecting a 16-bit parallel address bus to each chip might eliminate pins needed for chip testing either by the manufacturer or by the BVD software. In addition, there may not be enough space on the single-board backplane for address drivers and wires to all chip memories.

The suggestion outlined below has been made in order to simply but thoroughly test every connection in the single-board BVD. Recall that a connection in the decoder consists of a VLSI chip metric output driver, the wire it drives to another chip, and the corresponding input driver on the other chip. There are 5120 such connections in the single-board BVD. By placing a two-input multiplexor at the input to each one of the four 12-bit state-metric registers in each butterfly, the registers may be chained together during a test mode. The four multiplexors together require only about the same number of transistors as one of the 10 full-adders or one of the 60 flip-flops in a butterfly. The test mode could be used to check metric wires between chips after all chips have been successfully tested. Using the above multiplexors, a bit pattern may be loaded serially into state metric registers. If the decoder then processes 6 all-zero symbols, the pattern will be sent along metric wires between chips and into butterflies. The contents of all state-metric registers may then be read out bit-serially and used to identify bad metric wires directly. Traceback read operations are avoided, thereby reducing circuitry on the board, software, chip pins, and wiring of memory addresses to all VLSI chips.

Naturally, a progressive testing method for the decoder should be used in which successive tests depend upon the decoder's passing all previous tests. The datapath for symbols should be checked first. Then simple checks could be run on one chip to verify that control signals are working. Next, each chip, including the internal traceback memory, would be tested individually. This might be accomplished by using on-chip self-test circuitry. Each chip could be verified as a complete (9,1/6) Viterbi decoder. Then metric wires between chips could be checked using the procedure described above. Finally, the entire single-board decoder would be tested by comparing measured bit error rates for various noise levels against precomputed values from software simulations [3].

References

[1] J. Statman, G. Zimmerman, F. Pollara, and O. Collins, "A Long Constraint Length VLSI Viterbi Decoder for the DSN," *TDA Progress Report 42-95*, vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 134–142, November 15, 1988.

[2] O. Collins, F. Pollara, S. Dolinar, and J. Statman, "Wiring Viterbi Decoders (Splitting deBruijn Graphs)," *TDA Progress Report 42-96*, vol. October–December 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 93–103, February 15, 1989.

[3] I. M. Onyszchuk, "Coding Gains and Error Rates from the Big Viterbi Decoder," *TDA Progress Report 42-106*, vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 170–174, August 15, 1991.

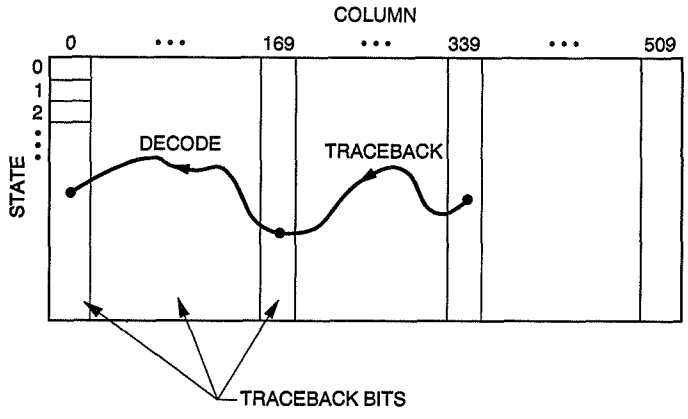


Fig. 1. Viterbi decoder traceback memory.

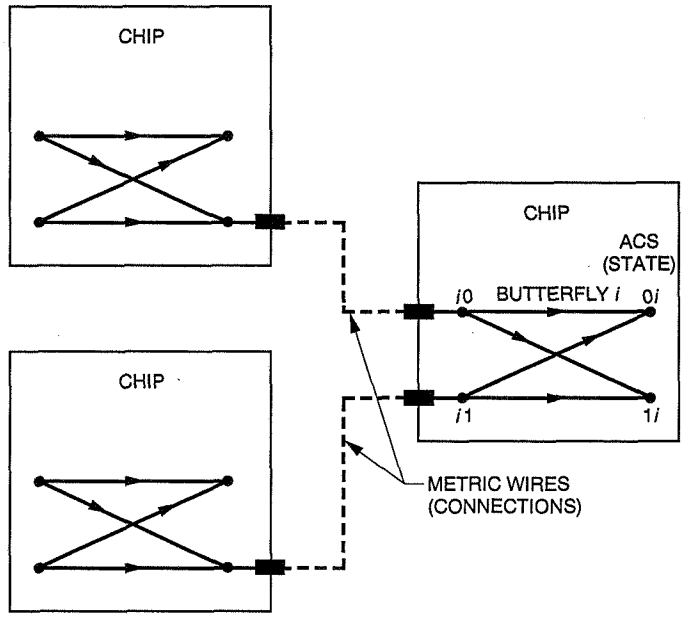


Fig. 2. Uncertainty of error location.

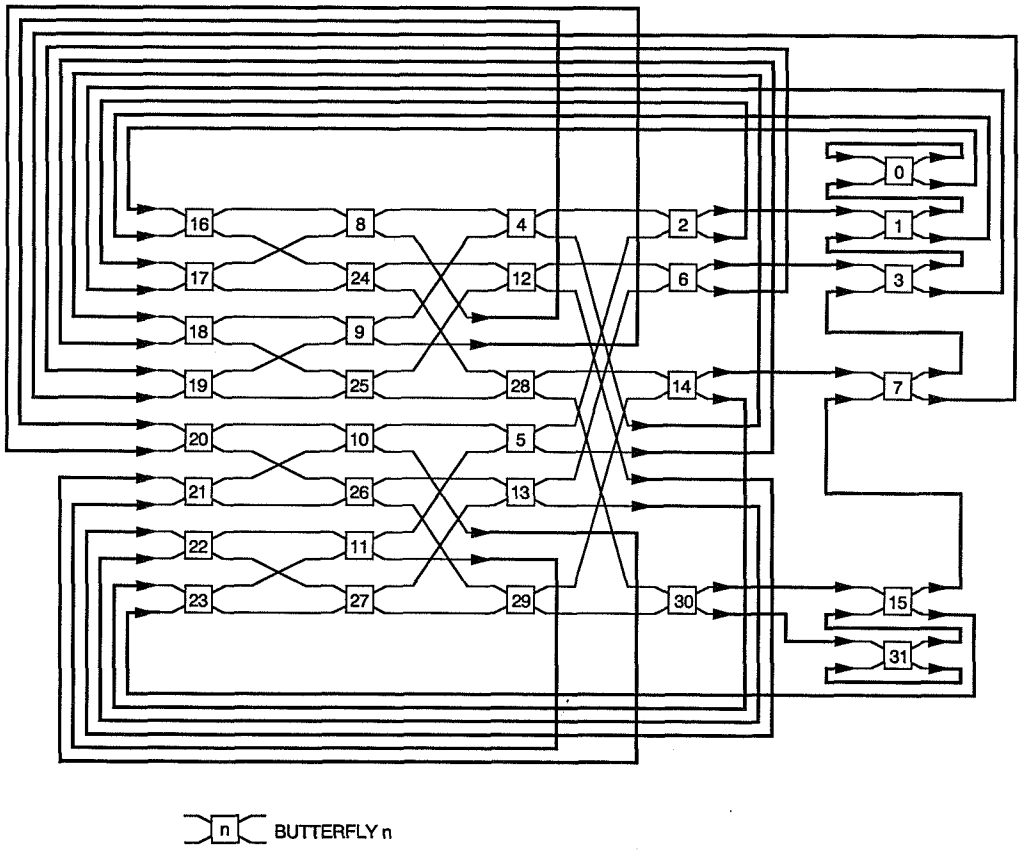


Fig. 3. Thirty-two butterflies connected to form the Add-Compare-Select (ACS) section of a constraint-length 7 Viterbi decoder.

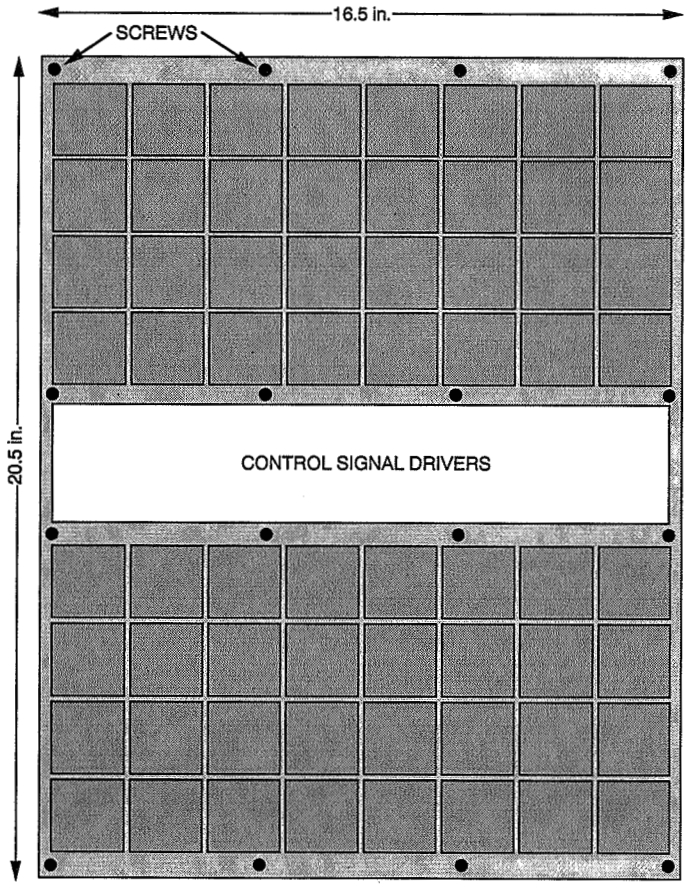


Fig. 4. A possible layout for the single-board Big Viterbi Decoder.

Frame Error Rate of the NASA Concatenated Coding System

L. Swanson and K.-M. Cheung
Communications Systems Research Section

For a concatenated coding system, the effect of interleaving depth on interleaved frame-error rate is compiled by simulation. Comparisons are made for three different inner convolutional codes; all three systems have a (255,223) 8-bit Reed-Solomon code as an outer code. For all the inner convolutional codes tested, an interleaving depth smaller than 4 causes substantial loss.

I. Introduction

Communication between spacecraft and the DSN has, for many years, included error-correcting codes, also known as channel codes. A JPL, NASA, and Consultative Committee for Space Data Standards (CCSDS) standard for channel coding is the concatenated code shown in Fig. 1. On the spacecraft, data are first encoded with a Reed-Solomon code, which adds 32 parity bytes to each block of 223 (8-bit) information bytes, and then the output of that block is encoded by a constraint length 7, rate 1/2 convolutional code. On the ground, the process is reversed: a convolutional decoder using the Viterbi algorithm is followed by a Reed-Solomon decoder. Presently, the DSN performs the convolutional decoding and passes the convolutionally decoded data to the project for Reed-Solomon decoding; in the future (beginning with Mars Observer) the DSN will do Reed-Solomon decoding. This system was first used by NASA during Voyager 2's encounters with Uranus and Neptune to protect compressed imaging data, and is typically used by a project to protect compressed data, because it is much more efficient than a convolutional code alone at obtaining low (10^{-6}) bit error rates.

There are several reasons for this concatenated system. The convolutional code with a Viterbi decoder is very good at dealing with extremely noisy data, while the Reed-Solomon code makes a medium channel extremely good, and identifies almost all of its own errors. The channel errors which are not corrected by the convolutional code tend to occur in bursts. The Reed-Solomon code is good at correcting short bursts of errors because it considers any errors in one byte as one error.

Because the error bursts in Viterbi-decoded data are often longer than one byte, the Reed-Solomon codewords are block interleaved to disperse error bursts over several codewords. Interleaving is best described by Fig. 2. The data stream is read into the array horizontally, and then block encoded vertically. Then the data stream is read out horizontally, as it was originally read in, and goes to the convolutional encoder. This means that a byte from one Reed-Solomon codeword is convolutionally encoded next to a byte from a different Reed-Solomon codeword. Because of this, bursts of Viterbi decoder errors cause a small number of errors in each of several Reed-Solomon words instead of causing a large number of errors in one Reed-

Solomon word. The horizontal dimension of the array in Fig. 2 is the interleaving depth, and the vertical dimension is the Reed-Solomon codeword size (255 for these uses). The whole array constitutes one interleaved frame. The interleaving depth for the CCSDS standards is five.

Galileo is using an experimental constraint length 15 convolutional code as the inner code in its concatenated system, and its Reed-Solomon code is interleaved to depth two. The constraint length 15 code was chosen to increase performance substantially compared to the standard constraint length 7 code. But the error bursts from a constraint length 15 code are typically about twice as long as those from a constraint length 7 code. On the other hand, the fact that (for historical reasons involving the length of data blocks) Galileo's Reed-Solomon code is interleaved to depth two means that it cannot handle long bursts of errors as well as it would with a larger interleaving depth.

An earlier article [1] quantified the loss in coding gain from the use of a small interleaving depth for outer Reed-Solomon codes concatenated with long constraint-length convolutional inner codes. That article measured performance in terms of average bit error rates. For some applications, bit error rate may not be the right measure of performance. Because bytes from one Reed-Solomon codeword are interleaved in the data with bytes from the other codewords in the same interleaved frame, it could be necessary to declare an entire interleaved frame in error whenever a decoded word in the frame has a detected error. The effect of the interleaving depth on frame errors is not immediately clear. Once codewords are interleaved "enough," an additional increase in interleaving depth will not cause much decrease in Reed-Solomon word error rate, but frame error rates will rise because the frames are longer. One would expect the frame error rate, as a function of interleaving depth, to fall for a while, and then to rise towards 1 as the interleaving depth goes to infinity; it is desirable to know where this turnaround takes place. Thus, the frame error rate statistics have been compiled in this article. A frame is declared in error if any Reed-Solomon word in the frame has more than the 16 byte errors which the code can correct. (In practice, a frame is not declared in error unless the Reed-Solomon decoder detects at least one word error in the frame, but almost every Reed-Solomon decoder error is detected, see [2].)

II. Simulation Results and Analysis

A study was conducted by simulation of a Gaussian channel at various signal-to-noise ratios using the Little Viterbi Decoder, a 100-bit-per-second decoder developed as a research tool by the Coding and Modulation Work Unit for the DSN Advanced Systems Program. The plots for various interleaving depth of frame error rates as a function of bit signal-to-noise ratio, E_b/N_0, are shown in Fig. 3 for the constraint length 7, rate 1/2 code. Figure 4 shows the same thing with Galileo's constraint length 15, rate 1/4 code used as the inner code. Figure 5 uses the constraint length 15, rate 1/6 code described in [3], which will be used by the Comet Rendezvous/Asteroid Flyby (CRAF) and Cassini missions. As can be seen in the graphs, interleaving depths between 4 and 8 give the best performance. The graphs do not have as much data, and especially as much low error-rate data, as one might like, because of the great length of simulation time required to tell anything about 10^{-5} error rates.

Acknowledgment

The authors would like to thank Sam Dolinar of Section 331 for his helpful suggestions.

References

[1] K.-M. Cheung and S. J. Dolinar, "Performance of Galileo's Concatenated Codes with Nonideal Interleaving," *TDA Progress Report 42-95*, vol. July–September 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 148–152, November 15, 1988.

[2] R. J. McEliece and L. Swanson, "On the Decoder Error Probability for Reed-Solomon Codes," *IEEE Transactions on Information Theory*, vol. IT-32, no. 5, pp. 701–703, September 1986.

[3] J. H. Yuen and Q. D. Vo, "In Search of a 2-dB Coding Gain," *TDA Progress Report 42-83*, vol. July–September 1985, Jet Propulsion Laboratory, Pasadena, California, November 15, 1985.

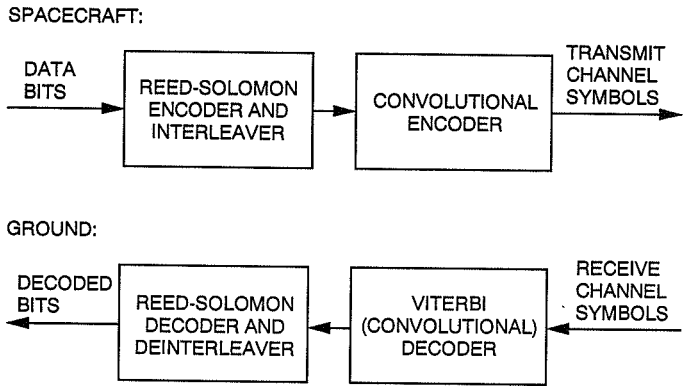


Fig. 1. Concatenated coding system.

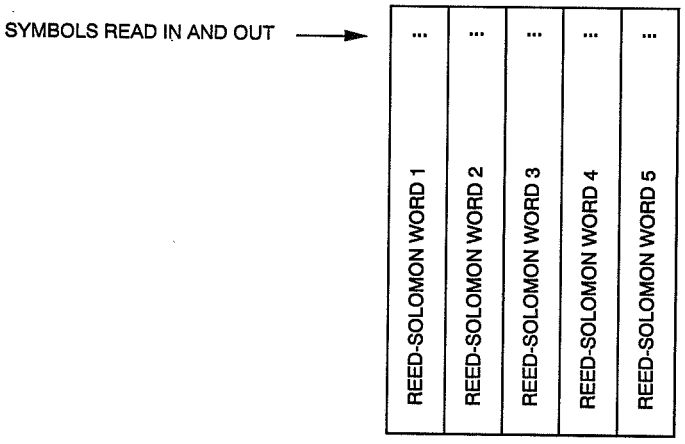


Fig. 2. Block interleaving (depth 5).

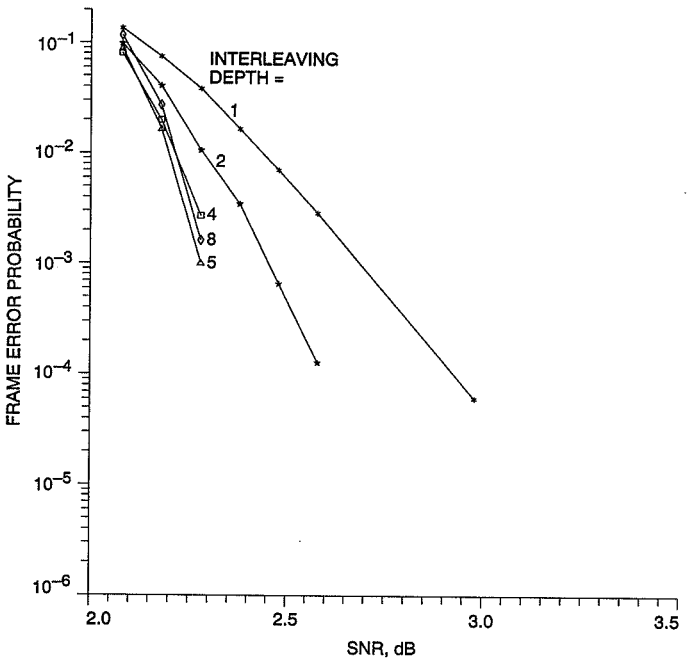


Fig. 3. Frame error probability of the concatenated coding scheme with (7,1/2) convolutional code with (255,223) Reed-Solomon code.

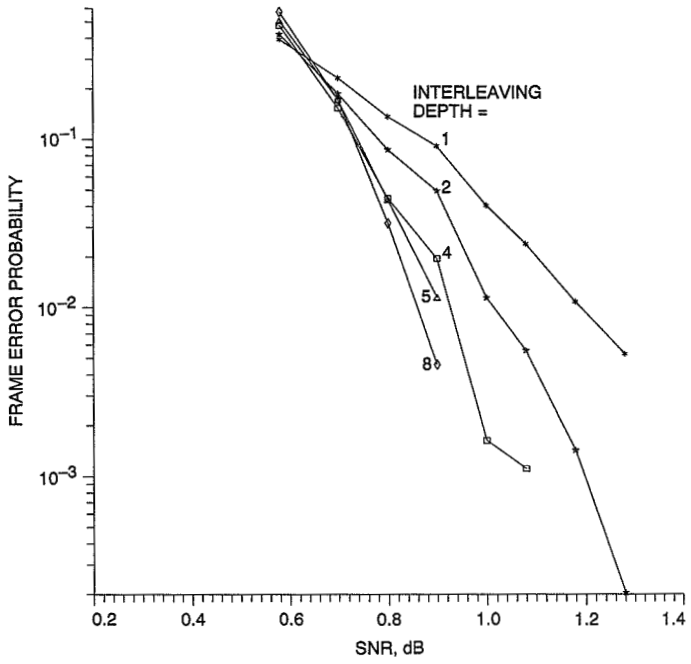


Fig. 4. Frame error probability of the concatenated coding scheme with (15,1/4) convolutional code with (255,223) Reed-Solomon code.

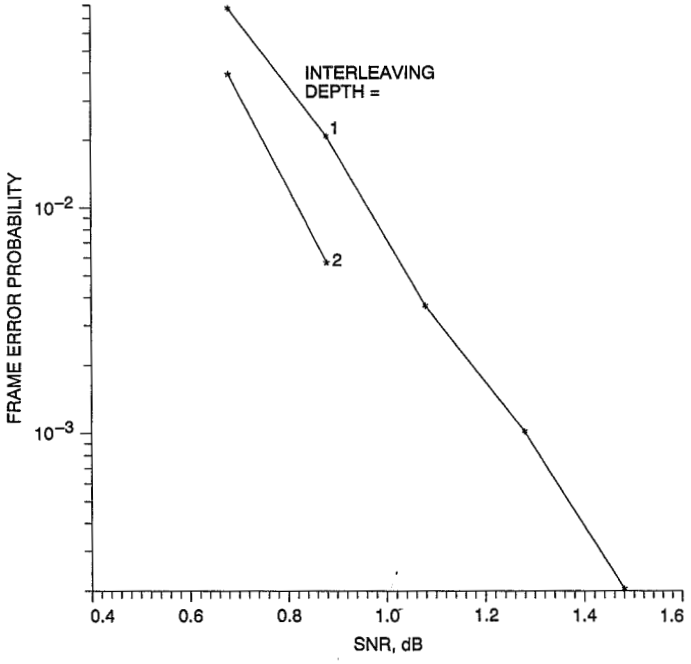


Fig. 5. Frame error probability of the concatenated coding scheme with (15,1/6) convolutional code with (255,223) Reed-Solomon code.

Initial Pointing Calibrations for the DSS 13 34-Meter Beam-Waveguide Antenna

L. S. Alvarez
Ground Antennas and Facilities Engineering Section

The beam pointing of the new DSS 13 beam-waveguide antenna at the Goldstone Venus site was calibrated during the postconstruction performance testing period from July 1990 through January 1991. The pointing calibrations were based on errors measured on radio sources at both the Cassegrain and centerline beam-waveguide focal points. The blind pointing performance goal of 5.0 mdeg, 3-sigma at Ka-band (32 GHz) was demonstrated to be met for low (<10 mph) wind conditions.

I. Introduction

This article describes the beam pointing for the new DSS 13 34-m beam-waveguide (BWG) antenna at the Goldstone Venus site from July 1990 through January 1991. An outline drawing of the DSS 13 antenna is shown in Fig. 1. During the calibration period, efficiency and pointing performance were characterized at X- and Ka-band frequencies (8450 MHz and 32 GHz, respectively) at both the f1 (Cassegrain) and f3 (centerline BWG) focal points illustrated in Fig. 1.

The objectives of the DSS 13 pointing calibrations were (1) to meet the specifications stated in the project Functional Requirements Document, and (2) to provide a systematic pointing-error model sufficient to carry out efficiency measurements. The beam-pointing accuracy required in less than 10-mph wind is 8 mdeg for X-band and 5 mdeg for Ka-band. These are 3-sigma blind-pointing specifications over the whole sky, with the pointing error defined to be the root sum square of the simultaneous errors sensed in the elevation and cross-elevation axes. It was decided that the 5-mdeg specification would be the goal for antenna pointing at both frequencies.

II. Pointing Calibration Overview

The primary objective of the antenna beam-pointing calibrations is to generate a pointing-error-correction model to be loaded into the antenna pointing system so that systematic pointing errors can be compensated for during tracking operations. The basic iterative procedure is to first measure pointing errors on a collection of radio sources adequately distributed over the sky. The negatives of the measured errors, termed corrections, are then fitted by the method of least squares to a set of pointing-error model terms. The resultant coefficients are loaded into the antenna control subsystem (ACS) computer, which then generates real-time pointing corrections as a function of antenna azimuth and elevation position. The process is repeated until suitable and predictable pointing perfor-

mance is attained. A detailed summary of the DSS 13 f1 and f3 pointing calibration methods employed and final results obtained are given below.

III. Beam-Pointing Measurements

A boresighting technique developed in 1989–1990 by R. L. Riggs of the Ground Antennas and Facilities Engineering Section was used to accurately point the antenna at the radio sources used in the calibration process. The method used was a "seven-point" boresight technique. This method moved the antenna sequentially along the cross-elevation (xel) and elevation (el) axes, both on and off the source. In each axis, the antenna was commanded to move off-source by 10 half-power (one-sided) beamwidths, one half-power beamwidth at the "3-dB" point, approximately 0.576 half-power beamwidths at the "1-dB" point, and on-source. Then, similar position offsets are executed for the other side. Noise-temperature measurements obtained at each offset are normalized with respect to a baseline generated by the off-source measurements.

The adjusted data points for a single axis are fit to a linearized exponential from which computed coefficients yield estimates of pointing error, half-power beamwidth, and peak temperature. Pointing measurements resulting from each axis scan are called scan errors. A pair of scans (one xel and one el) are considered to be one measurement point. The seven pointing offsets for each new scan are corrected for pointing errors found from the previous scans during the track. This is accomplished by centering those offsets about the running sum of the scan pointing corrections and then sending them as position offsets to the antenna via the ACS. The scan corrections are just the negatives of the scan errors. During the measurements, the principal outputs of the calibration software are an efficiency file and a pointing file. The efficiency file consists of time, azimuth, elevation, and xel and el estimates of half-power beamwidth, peak source temperature, and scan pointing error. The pointing file consists of azimuth, elevation, and accumulated pointing corrections in the xel and el axes. Hereafter, the accumulated pointing corrections are referred to simply as pointing corrections.

During the early period of the measurements described in this article, the boresight program resided in a PC and noise-temperature measurements were entered by an operator. Pointing the antenna was executed from the local control display (LCD). The basic calibration test configuration used is shown in Fig. 2. During 1990–1991, S. R. Stewart of the Ground Antennas and Facilities Engineering Section integrated improved boresight and radiometer system software into a single PC-based module that both controls the radiometer system hardware and directly sends offsets to the antenna servos through the ACS maintainance port. A full (two-axis) boresight now takes approximately 7 min to complete. Further discussion on the boresight methodology and calibration instrumentation may be found in [1,2] and in the DSS 13 BWG antenna Phase 1 Final Report.[1]

IV. Systematic Pointing-Error Modeling

A. Model Generation

Pointing corrections are used to assess performance and also to create or update the systematic pointing-error models. The plot of these numbers versus antenna position is theoretically the error path that the antenna beam would follow if no external corrections at all were loaded into the ACS after each boresight scan. This residual error path then represents just the imperfections of the present error model in use and other antenna boresight corrections resident in the ACS, as well as random error contributions. Its magnitude is thus a measure of the antenna blind-pointing performance. Assuming environmental effects to be minimal and other ACS pointing compensation (e.g., squint, refraction) to be accurate, one may attribute the corrections to the inadequacy of systematic error modeling. The model-building process then proceeds by adding the corrections to those corrections generated by the a priori model used in that particular track. These new values are called total corrections and are used to generate an updated error model.

The Cassegrain-focus systematic pointing-error model for an elevation-over-azimuth mount antenna is shown in Table 1. As shown, the systematic pointing-error list includes such contributors as encoder fixed offsets, azimuth plane tilt, gravitational flexure, and fixed source coordinate errors. A more thorough discussion of the parameters as presented in a 1986 seminar[2] may be found in [3]. PC-based pointing-model software utilized at DSS 13 enables the merging of correction files from multiple star tracks into one total correction file. A second program

[1] M. J. Britcliffe, L. S. Alvarez, D. A. Bathker, P. W. Cramer, T. Y. Otoshi, D. J. Rochblatt, B. L. Seidel, S. D. Slobin, S. R. Stewart, W. Veruttipong, and G. E. Wood, *DSS 13 Beam Waveguide Antenna Project; Phase 1 Final Report*, JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

[2] R. L. Riggs, "Antenna Pointing Angle Corrections," DSN Antenna Seminar, Videotapes 49-54 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 1986.

then performs a least-squares fit of the data to any combination of parameters selected from Table 1. The output is a new set of model coefficients (P values in Table 1) that can be manually entered into the ACS. The ACS will then compute pointing corrections from those parameters for any combination of azimuth and elevation. In general, the pointing correction modeling of DSS 13 followed the techniques and guidelines given in the 1986 seminar[3] and in [4].

B. Measurement and Parameter Uncertainties

The modeling process encompasses much analysis and engineering judgment. It involves data archiving and editing correction files from the star tracks. Pointing errors generated by nonsystematic sources such as wind, thermal, refraction, mechanical, and radiometer anomalies must be recognized and edited out as much as possible. The total corrections should be repeatable and uncontaminated by these effects. In the parameter estimation, it is assumed that the pointing data provided by the microwave calibration instrumentation are accurate and measurement uncertainty is small as compared with the magnitude of the data points.

Small calculated pointing errors may be due to random errors in the system noise-temperature measurements rather than actual antenna mispointing. As discussed in [2], the stability of the X-band front-end test package system noise temperature on the ground was very good, with no discernible changes greater than 0.02 K. Depending on the antenna elevation and the strength of the particular radio source being observed, the propagation of 0.01 to 0.1 deg K would result in an uncertainty in the pointing-error estimate of no more than a few tenths of a millidegree. Of greater concern is the on-source temperature measurement uncertainty introduced by the dynamics of the antenna structure and axis servos. In higher than average wind conditions (10–20 mph), the signal jitter (due to mechanical oscillations) observed during measurements on the side of the antenna beam is significantly increased. Through observations of the elevation axis encoder, the mechanical pointing errors in such conditions were seen to approach 2.0 mdeg peak-to-peak (1-sigma). This value is of the same order of magnitude as the preliminary analytical results reported in [5]. To minimize errors incurred in the estimates of the static pointing and efficiency variables (especially at the Ka-band frequency), system temperatures are integrated over a user-specified time interval [2], generally on the order of 3–10 sec. Consistent logging of the weather data enabled the flagging of wind-corrupted pointing corrections.

[3] Ibid.

The quality of a generated model is also dependent on the distribution of the measured beam-pointing errors. The linear dependence of the columns of the measurement-distribution matrix that forms the least-squares problem is strongly dependent on the sky distribution of the observation points [4]. To ensure a satisfactory matrix condition that leads to accurate estimation of all the parameters in Table 1, radio sources at different declinations must be tracked to ensure full ranges of both azimuth and elevation. Various numerical statistics in the modeling program allow assessment of this matrix condition as well as the performance of the parameter fit in terms of estimation residuals.

V. Pointing Calibrations at f1

All the f1 pointing-error modeling was based on X-band tracks. Four iterations of the Cassegrain-focus model were made during the first three months of the f1 calibrations. The final model was obtained during the week of post-holography X-band calibrations, with the antenna main reflector panels in their final adjustment condition. It was based on tracks of the radio sources DR21 and 3C274 (declinations 42.3 and 12.4 deg, respectively). Consecutive tracks on these declinations showed the most repeatability in the pointing, while data from other sources were degraded by high winds (greater than 15 mph) and antenna mechanical malfunctions. The total corrections for each beam axis along with the f1 model predictions are shown in Figs. 3 and 4. The azimuth corrections (which are applied to the azimuth axis servo) are computed from the cross-elevation corrections and are illustrated in Fig. 5. The model coefficients and their theoretical standard deviations derived from the least-squares covariance matrix are shown in Table 2.

The constant declination and hour angle error parameters $P6$ and $P21$ are not included because they were estimated with insignificant magnitude. Only $P2$ was estimated, since the simultaneous estimation of $P1$ and $P2$ resulted in large matrix condition. The elevation corrections imply a negative fixed encoder offset and large residual squint compensation as indicated by $P8$. The different paths for the rising and setting over the tracks shown in Fig. 3 are due to the azimuth dependence of the elevation correction required for the azimuth plane tilt modeled by the $P4$ and $P5$ terms. Theodolite measurements of the DSS 13 azimuth track indicated only a slight plane tilt of 2 mdeg, where, as seen in Table 2, the tilt coefficient is estimated to be almost 8 mdeg. The additional error magnitude was possibly due to an intermittent el-

evation position-encoder error.[4] An apparent motion of the elevation bearing relative to the encoder resulted in a hysteresis-like position-error signature relative to the direction of elevation motion. These errors were sensed as a function of azimuth during the star tracks. Since the f1 calibrations, the magnitude of the problem has apparently been reduced by redesign of the coupling between the antenna structure and the encoder.

Not many data points at azimuths greater than 300 deg were available to use in the modeling. As a result, the model's predictive capability in that region of the sky is degraded, as shown in Figs. 3–5. This also increased the overall standard deviations of the model estimation residuals. They were 2.0 mdeg in xel and 2.6 mdeg in el, resulting in a large root-sum-square standard deviation of 3.3 mdeg.

Figure 6 shows the azimuth–elevation plane trajectories of the star tracks over which the DSS 13 pointing performance will be illustrated. The data points are actual boresight measurement locations. Figure 7 illustrates the beam-pointing corrections resulting from a Ka-band track of 3C274 at f1, whose trajectory is shown in Fig. 6. The measurements were taken in low winds (5–9 mph) and clear skies. The Cassegrain-focus model described above was used during the track. The beam corrections are defined to be the root sum square of the elevation and cross-elevation corrections as computed by the calibration program. As illustrated, the f1 model predicts very well at the high elevations but is degraded at the lower end. Refraction uncertainties at the lower end degraded pointing and systematic error modeling during the entire f1 calibrations. As seen in Fig. 3, very few data points below 20 deg elevation were deemed reliable enough to be incorporated into the measurement distribution. The corrections in Fig. 5 are typical of those sensed during other star tracks at both lower and higher declinations in a favorable environment. The scan-to-scan errors sensed approximately every 10–15 min (obtained by differencing the sequential corrections in Fig. 7) are mostly on the order of 1 mdeg or less. The 3-sigma blind-pointing tolerance of 5.0 mdeg was demonstrated to be met at elevations greater than 17 deg.

VI. Beam-Waveguide Pointing Errors

The calibration of the f3 beam pointing of DSS 13 is unique in that it involves errors contributed by imperfections of the BWG system. Static mirror misalignments and deflections of the mirrors as a function of antenna orientation induce pointing errors that may be sensed at the f3 focus. The pointing calibration approach was to first estimate these errors during the mirror-alignment stage, and then, if possible, try to compensate for them with the existing Cassegrain-focus pointing model.

During the center-fed BWG mirror alignment, center-ray laser theodolite deflection measurements were obtained by M. J. Britcliffe of the Ground Antennas and Facilities Engineering Section. From the antenna pedestal room, a laser at f3 was focused onto a target 100 in. below the antenna hub through all the BWG flat-plate alignment mirrors. The optical measurement configuration is shown in Fig. 8. The antenna was then moved in azimuth and elevation, and the laser focus was sketched on the target. A static misalignment of a mirror or feed in the pedestal room below the antenna track will result in an azimuth-dependent circular trace relative to the mirror set angles of azimuth = 340 deg and elevation = 45 deg. In elevation movement, the deflections of the BWG mirror that is mounted next to the antenna hub (M1 in Fig. 8) cause a change in the alignment of the BWG system and result in pointing errors above and below the 45-deg set angle.

The measured displaced center ray-traces were treated as a translated f1 feed in a Cassegrain antenna, from which crude first-order pointing-error predictions were then computed. This modeling approach ignored any equivalent Cassegrain f1 feed rotations resulting from an angled center ray impinging on the target. The cross-elevation beam errors were estimated to vary as a function of cos (el) and sin (az–340 deg). The expected variation of the elevation errors was with respect to cot (el) and cos (az–340 deg). Constant beam shifts were also expected in each axis. Expanding the sin (az–340 deg) and cos (az–340 deg) terms and reviewing the xel- and el-basis terms in Table 1 indicates that the parameters $P1$, $P2$, $P4$, $P5$, $P6$, $P7$, and $P9$ would accommodate estimation of the new pointing errors introduced by the centerline BWG system. Note that the $P6$ term typically deals with source declination error; however, its functional form aids in accommodating the expected BWG mirror-misalignment effects. The only term in the above expansion not available in the current xel error model is a cos (az) function.

VII. Pointing Calibrations at f3

During the X-band calibrations at f3, the f1 pointing model discussed earlier was used. The residual pointing errors remaining at f1 are shown in Fig. 7 and are then an uncertainty baseline for quantification of the new errors sensed at f3. The f3 pointing corrections in el and xel measured on a 3C84 track (declination 41.5 deg) are

[4] Britcliffe et al., op. cit.

shown versus elevation in Figs. 9 and 10. As indicated, the deflections of the first (and to a lesser extent, second) BWG mirror with elevation change produces beam errors at f3, primarily in the xel axis. An apparent discontinuity in the azimuth bearing rack and encoder gearing interface produces a jump in the cross-elevation corrections at approximately 62 deg azimuth. From the trajectory of source 3C84 shown in Fig. 6, this corresponds to approximately 40 deg elevation for clockwise azimuth travel and 73.5 deg elevation for counterclockwise motion. The former clockwise crossing consistently yielded a larger shift in azimuth (of roughly 9.5 mdeg) and was first observed during the f1 calibrations. A shift in the elevation corrections is also discernible at roughly 41 deg elevation; however, it was not found to be repeatable.

Of the measurement sets obtained during the week of X-band star tracks, only the 3C84 track shown and a full 3C274 track were deemed suitable for calibration purposes. High winds and a malfunctioning subreflector servo positioner contaminated the remainder of the pointing data. The former tracks were added to the f1 pointing-model corrections to yield total antenna corrections. After editing, they were fitted to the parameters $P2$–$P9$, yielding the f3 model presented in Table 3. A comparison of the coefficients above with the f1 model indicates that most of the BWG pointing deflections were picked up by the $P2$ through $P5$ terms and the additional $P6$ term.

There was very little change in the elevation error coefficients. The f3 model fit had residual standard deviations of 2.23 mdeg in xel and 1.91 mdeg in el. These residuals are defined to be the differences between the total el and xel corrections and the computed model predictions for each track, and are shown in Figs. 11 and 12. Both residual signatures in each axis are similar, with the exception of the discontinuities at about 40 deg elevation for the 3C84 track. The residuals also indicate that larger than desirable model uncertainties remain in both axes at the lower elevations, as well as at the meridian crossings.

This model was used in one more X-band track and the subsequent Ka-band calibrations. The el and xel corrections for the 3C123 X-band and two combined 3C84 Ka-band tracks, whose trajectories are illustrated in Fig. 6, are shown in Figs. 13 and 14. These postcalibration measurements were taken in winds less than 10 mph and were typical of the tracks at f3 in a favorable environment. The Ka-band measurements, in general, tended to be more scattered. The pointing-correction signatures remaining for both tracks shown are similar in elevation, thus suggesting no azimuth dependence. They also still resemble the general signature of the source 3C84 elevation corrections needed before the f3 calibration shown in Fig. 9. The cross-elevation residual pointing adjustments in Fig. 14 differ at low elevation and at the meridian crossings. The shifted 3C84 corrections below 40 deg elevation demonstrate the inability of the pointing model in Table 1 to accommodate the large clockwise azimuth discontinuity discussed above. The cause for the divergence of the 3C84 (northern meridian crossing) and 3C123 (southern meridian crossing) corrections in Fig. 14 is still not known.

Beam-pointing corrections computed from the measurements shown in Figs. 13 and 14 are illustrated in Fig. 15. The majority of the scan-to-scan pointing errors sensed for each track are less than or equal to 1.5 mdeg. The larger accumulated beam corrections, which were analyzed in the orthogonal axes above at the higher and lower elevations, were predicted by the f3 pointing-model estimation residuals. Efforts to resolve and compensate for the remaining f3 pointing errors are ongoing and may lead to augmenting the f1 Cassegrain-focus pointing model with extra terms. However, as illustrated in Fig. 15, the model terms shown in Table 1 were capable of meeting the DSS 13 f3 blind-pointing requirement of 5 mdeg at elevations greater than 25 deg.

VIII. Summary and Concluding Remarks

The methods employed and the final results obtained for the initial pointing calibrations of the DSS 13 BWG antenna from July 1990 through January 1991 have been presented. The primary objective for this first phase of the DSS 13 antenna project was to meet a 5.0-mdeg blind-pointing specification under low (less than 10 mph) wind conditions at both the f1 and f3 focal points. Pointing measurements (made simultaneously with efficiency measurements) demonstrated that this goal was achieved over a wide range of azimuths and at elevations greater than 17 deg at f1 and greater than 25 deg at f3. It was also shown that the current Cassegrain-focus systematic pointing-error model residing in the DSS 13 ACS was able to accommodate the pointing degradations introduced by the centerline BWG, at least to the level of accuracy required for the initial performance testing.

The DSS 13 antenna pointing calibration effort presented in this article is preliminary. A more thorough analytical modeling effort of the center-fed BWG system pointing corrections needs to be pursued. Repeatability of the remaining pointing correction residuals at f3 offers encouragement that further predictive performance may be achievable by augmenting the Cassegrain-focus systematic error model with new terms. With respect to station ca-

pabilities, the pointing will have to be continuously monitored and recalibrated through the many mechanical and microwave upgrades and experiments planned for this research and development antenna. For example, the beam pointing was observed to have shifted 30 mdeg in January 1991 after minor optical realignment of the ellipsoidal BWG mirror in the pedestal room. All the pointing-model software used in the preliminary calibrations is now resident at DSS 13. However, further training of DSS 13 personnel is recommended to ensure maintenance of the initial levels of accuracies achieved.

The characterization of pointing performance in higher wind conditions and at the lower elevation angles is still incomplete. On windy days (greater than 15 mph), it was observed that the DSS 13 beam pointing (measured roughly every 10 min) does not yet meet the 5.0-mdeg 3-sigma blind-pointing goal. The dynamic pointing errors, as sensed through the apparent noise-temperature jitter, were a problem at Ka-band and necessitated larger radiometer integration time intervals. More pointing measurements need to be taken in conjunction with the analytical investigations initiated in [5] in order to study and credibly quantify these wind-induced performance degradations. Likewise, more star tracking at the lower elevations is needed to classify the low-elevation pointing errors in terms of refraction uncertainty and/or mechanical deficiencies in the pointing model. The option for real-time surface weather inputs to the ACS refraction model is not yet implemented. This utility needs to be thoroughly tested and its performance measured against the default refraction model resident in the DSS 13 ACS.

Acknowledgments

The author thanks R. L. Riggs for the development of the boresighting program and pointing-model software used during the testing period discussed here. A team effort including M. J. Britcliffe, M. M. Franco, T. Y. Otoshi, S. D. Slobin, and S. R. Stewart provided the pointing and efficiency measurements. H. Ahlstrom provided many helpful discussions on the DSS 13 ACA, and M. J. Britcliffe provided the theodolite measurements necessary for the BWG pointing-error modeling. Lastly, the Goldstone DSS 13 personnel are gratefully acknowledged for their support throughout the entire performance-testing program.

References

[1] S. D. Slobin, "Efficiency Calibration of the DSS 13 34-Meter Diameter Beam Waveguide Antenna at 8.45 and 32 GHz," *TDA Progress Report 42-106* (this issue), vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 283–297, August 15, 1991.

[2] T. Y. Otoshi, S. R. Stewart, and M. M. Franco,"A Portable X-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation, Part I: Design and Ground Tests," *TDA Progress Report 42-103*, vol. July–September, 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 135–150, November 15, 1990.

[3] C. N. Guiar, F. L. Lansing, and R. Riggs, "Antenna Pointing Systematic Error Model Derivations," *TDA Progress Report 42-88*, vol. October–December 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 36–46, February 15, 1987.

[4] L. S. Alvarez, "An Analysis of the Least Squares Problem for the DSN Systematic Pointing Error Model," *TDA Progress Report 42-104*, vol. October–December 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 17–29, February 15, 1991.

[5] W. Gawronski and J. A. Mellstrom, "Modeling and Simulations of the DSS 13 Antenna Control System," *TDA Progress Report 42-106* (this issue), vol. April–June 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 205–248, August 15, 1991.

Table 1. Systematic pointing error sources and model terms

Error source	Model function	
	Cross-elevation error	Elevation error
Az[a] collimation	P_1 [d]	–
Az encoder fixed offset	P_2 cos (el)	–
Az/el skew	P_3 sin (el)	–
Az axis tilt	P_4 sin (el) cos (az)	$-P_4$ sin (az)
Az axis tilt	P_5 sin (el) sin (az)	P_5 cos (az)
Source dec[b]	P_6 sin (az)	P_6 sin (el) cos (az)
El[c] encoder fixed offset	–	P_7
Gravitational flexure	–	P_8 cos (el)
Residual refraction	–	P_9 cot (el)
Source HA[e]	P_{21} cos (dec)	–

[a] Az refers to azimuth angle.
[b] Dec refers to declination angle.
[c] El refers to elevation angle.
[d] P refers to parameter value.
[e] HA refers to hour angle.

Table 2. f1 systematic pointing-error model coefficients

Coefficient	Value, mdeg	Standard deviation, mdeg
$P2$	4.23	0.27
$P3$	10.85	0.22
$P4$	−7.70	0.12
$P5$	0.52	0.14
$P7$	−22.93	0.32
$P8$	−44.37	0.87
$P9$	−4.53	0.25

Table 3. f1 systematic pointing-error model coefficients

Coefficient	Value, mdeg	Standard deviation, mdeg
$P2$	16.23	0.36
$P3$	−5.36	0.27
$P4$	−9.43	0.15
$P5$	5.62	0.53
$P6$	−5.12	0.46
$P7$	−22.14	0.39
$P8$	−46.73	1.01
$P9$	−5.31	0.28

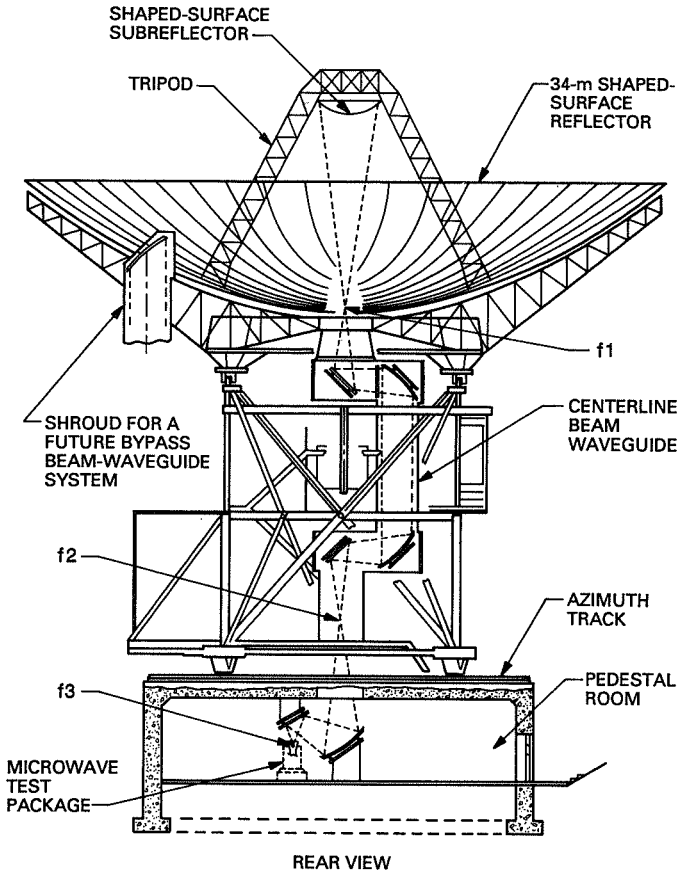


Fig. 1. Outline drawing of DSS 13 antenna.

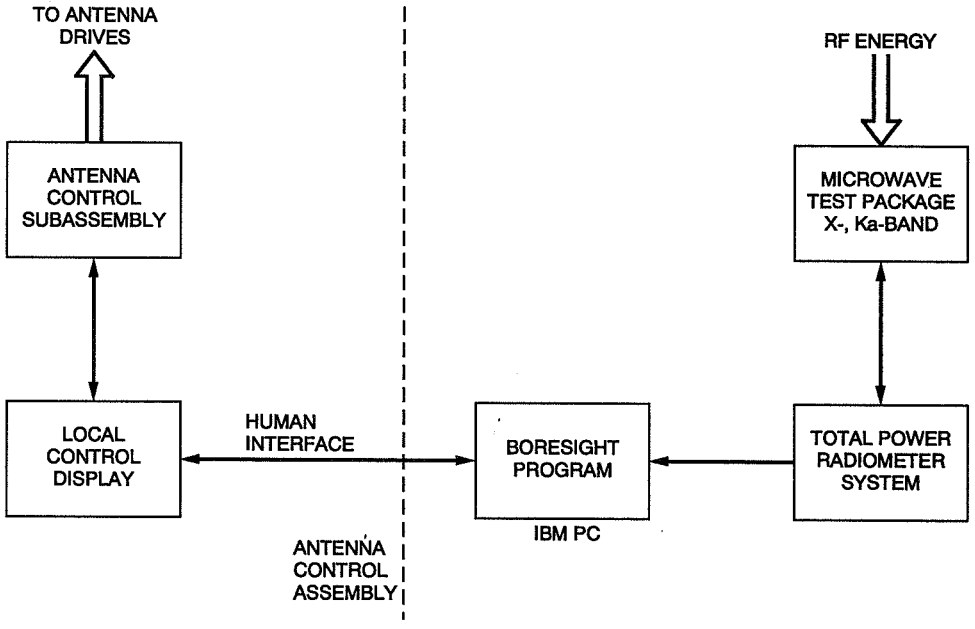


Fig. 2. Calibration instrumentation configuration.

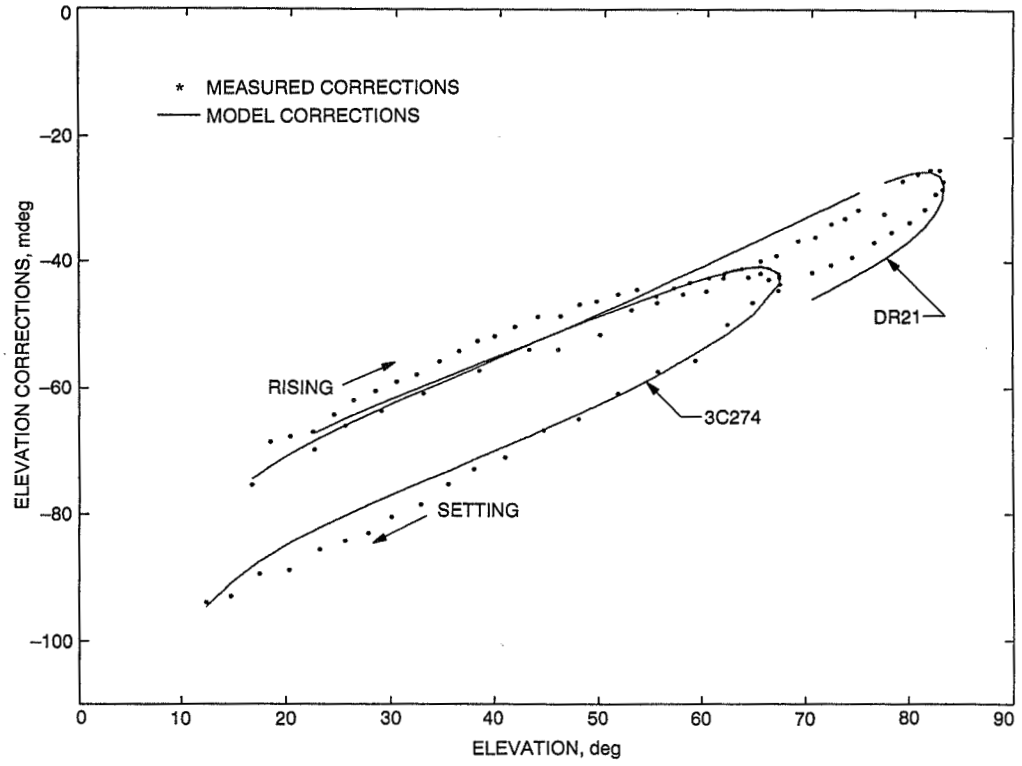


Fig. 3. f1 total elevation corrections versus f1 model predictions.

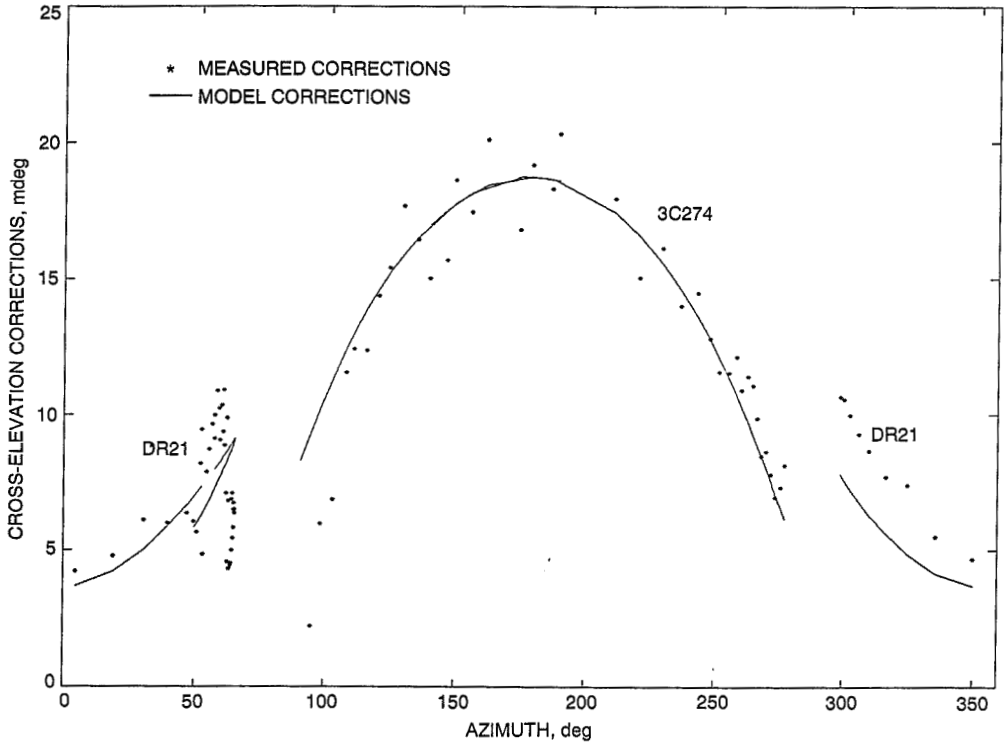


Fig. 4. f1 total cross-elevation corrections versus f1 model predictions.

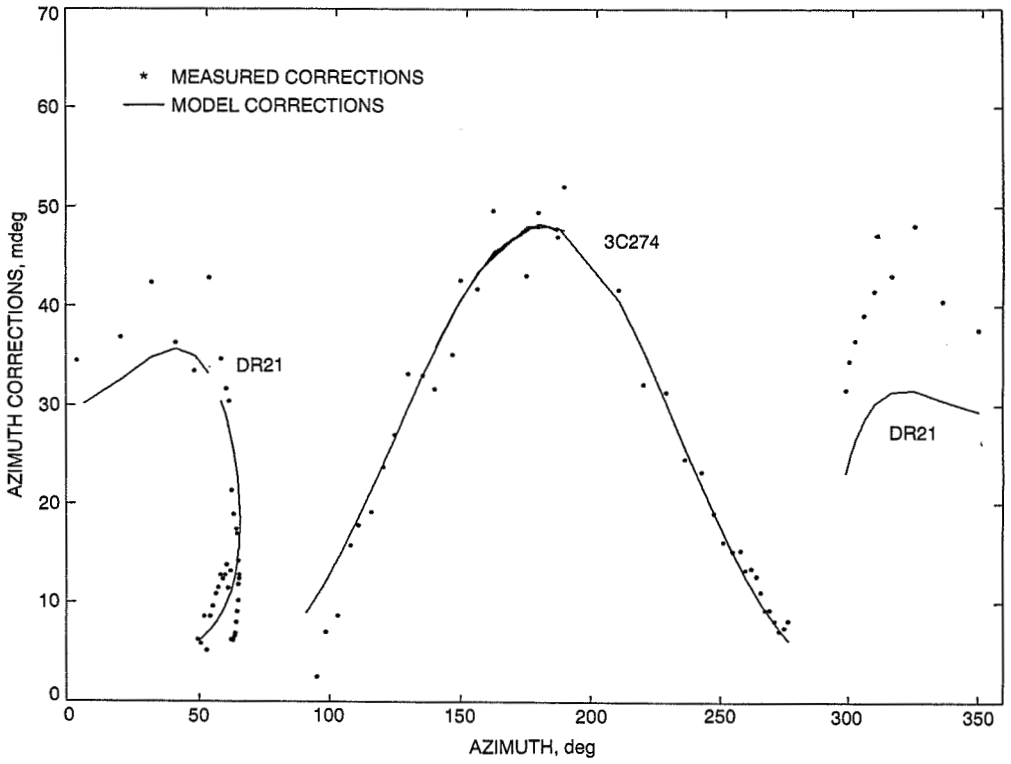


Fig. 5. f1 total azimuth corrections versus f1 model predictions.

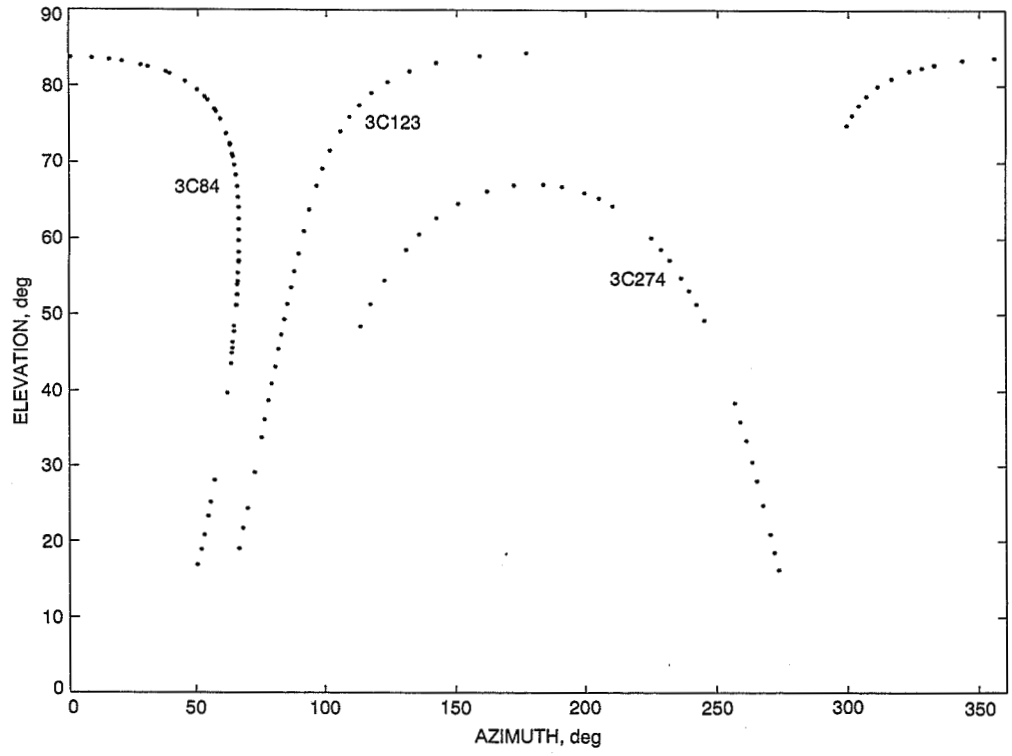


Fig. 6. Radio source trajectories for DSS 13 pointing performance assessment.

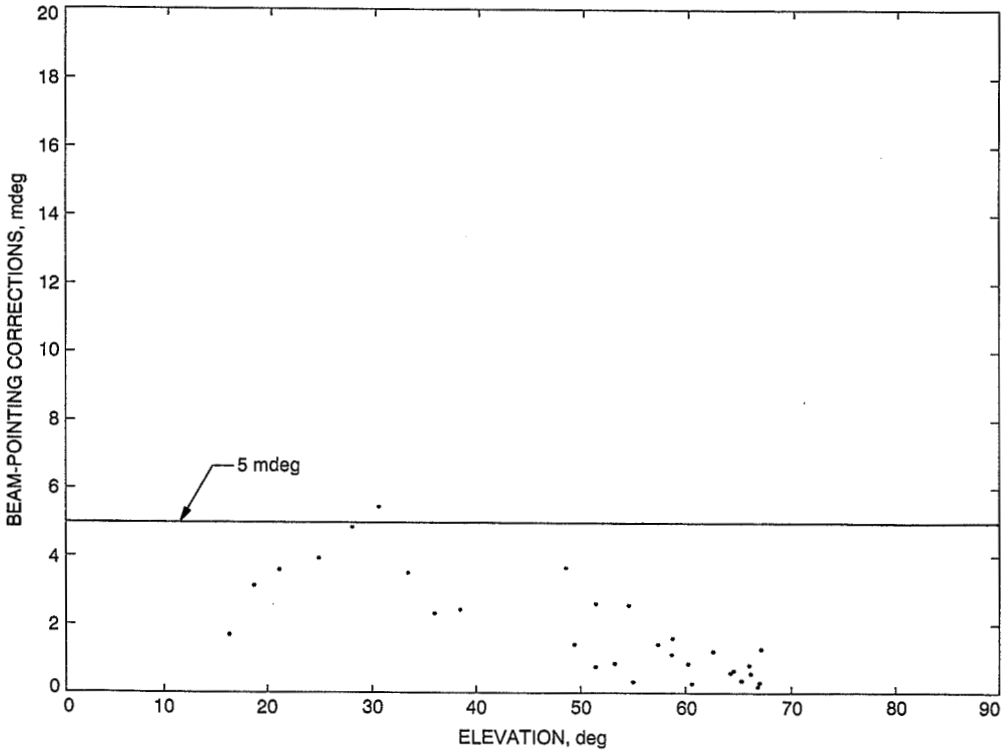


Fig. 7. Beam-pointing corrections for 3C274 track, DOY 288, 1990, Ka-band at f1.

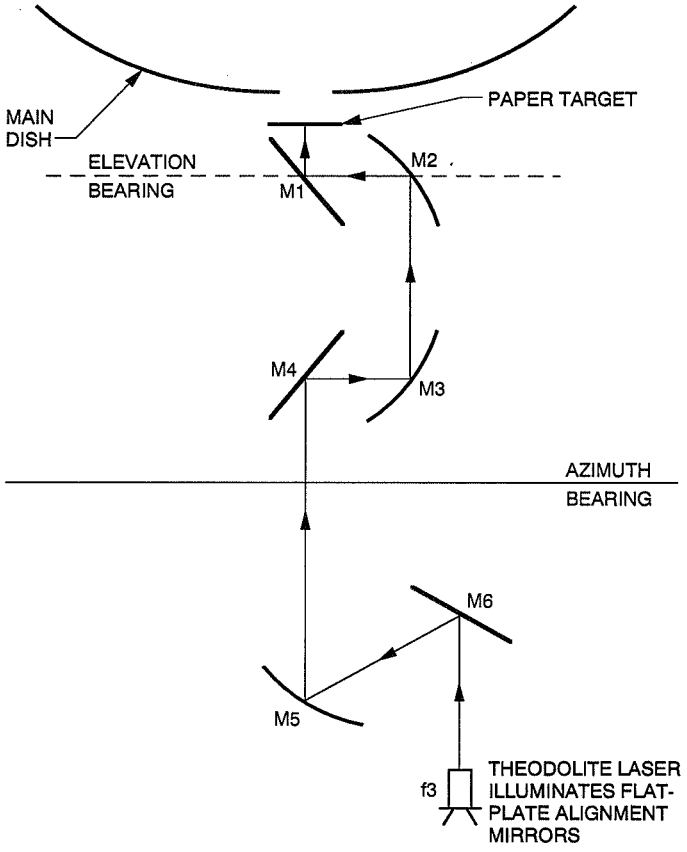


Fig. 8. Theodolite laser experiment configuration.

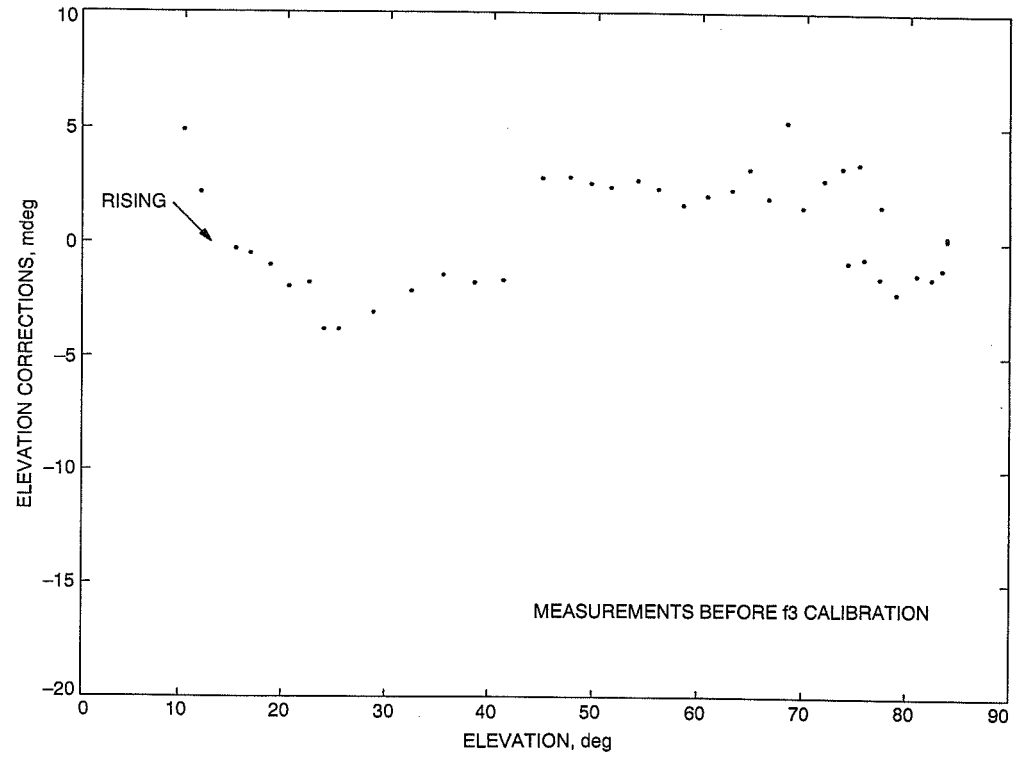


Fig. 9. Elevation pointing corrections for 3C84 track, DOY 311, 1990, X-band at f3.

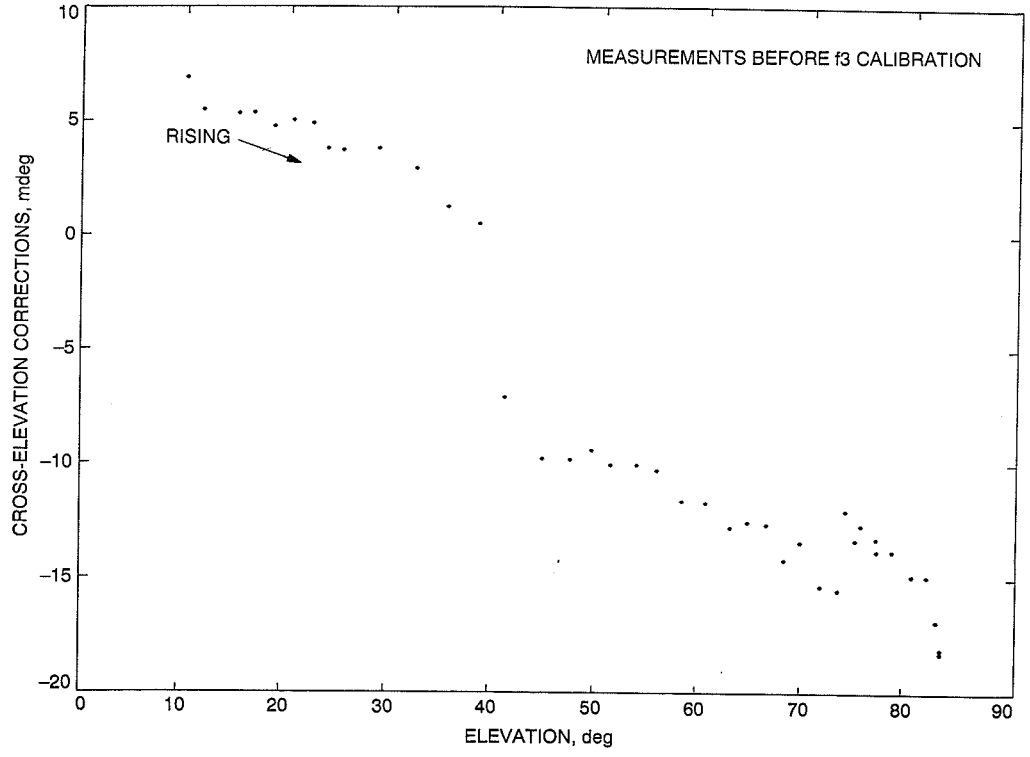


Fig. 10. Cross-elevation pointing corrections for 3C84 track, DOY 311, 1990, X-band at f3.

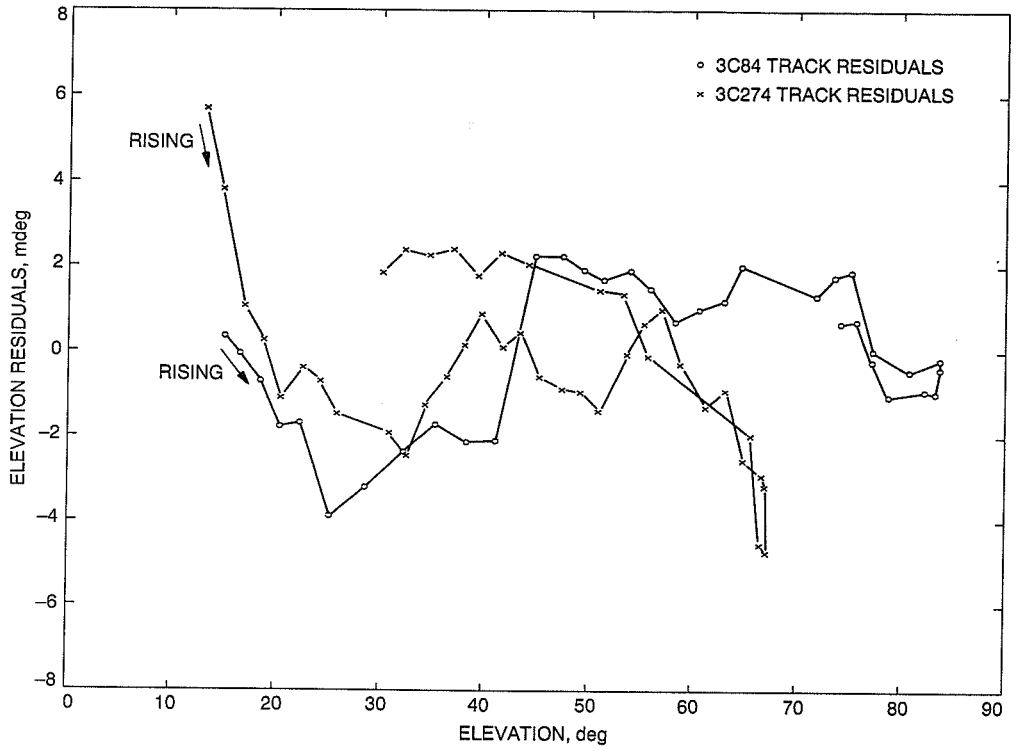


Fig. 11. f3 pointing model elevation residuals.

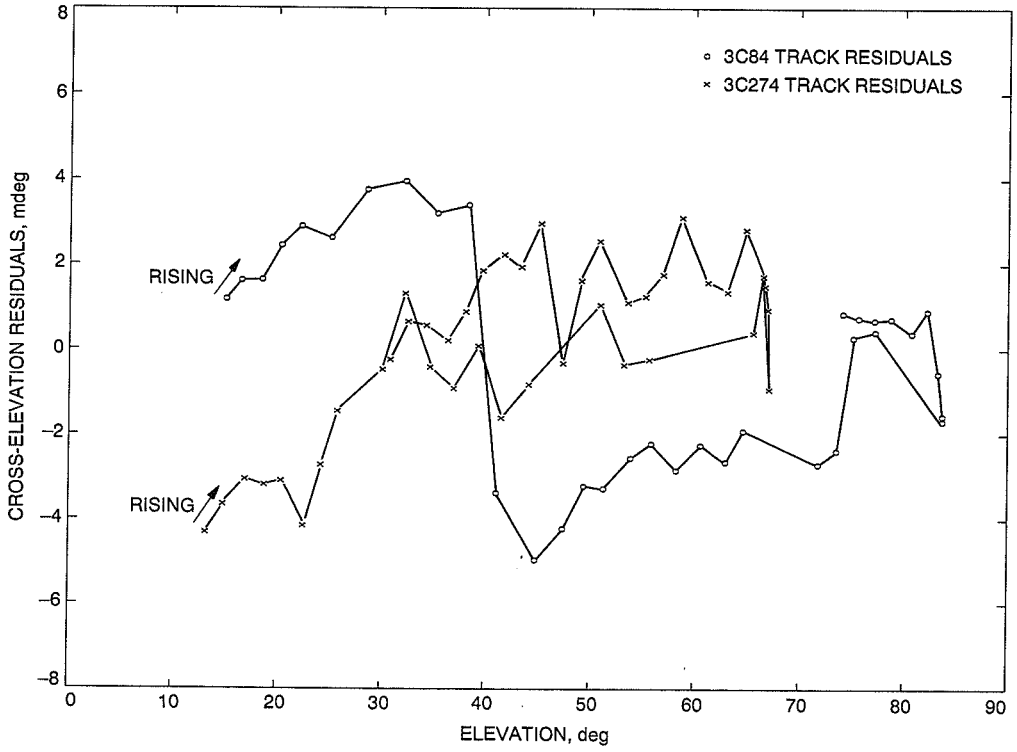


Fig. 12. f3 pointing model cross-elevation residuals.

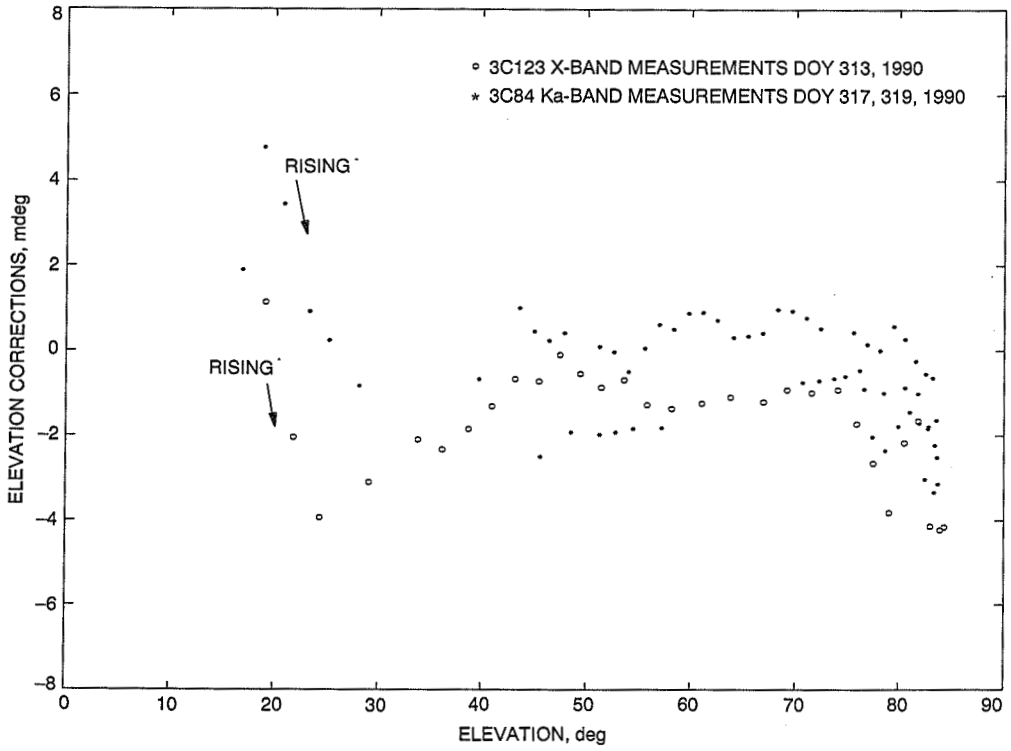


Fig. 13. Elevation pointing corrections for 3C123 and 3C84 tracks at f3.

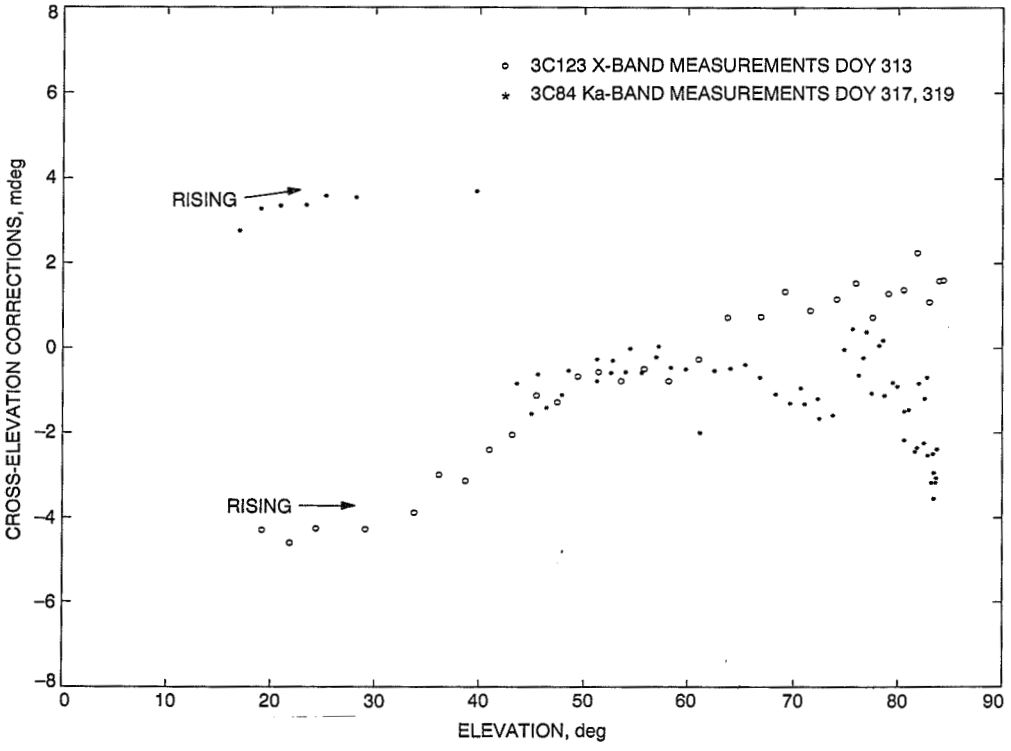


Fig. 14. Cross-elevation corrections for 3C123 and 3C84 tracks at f3.

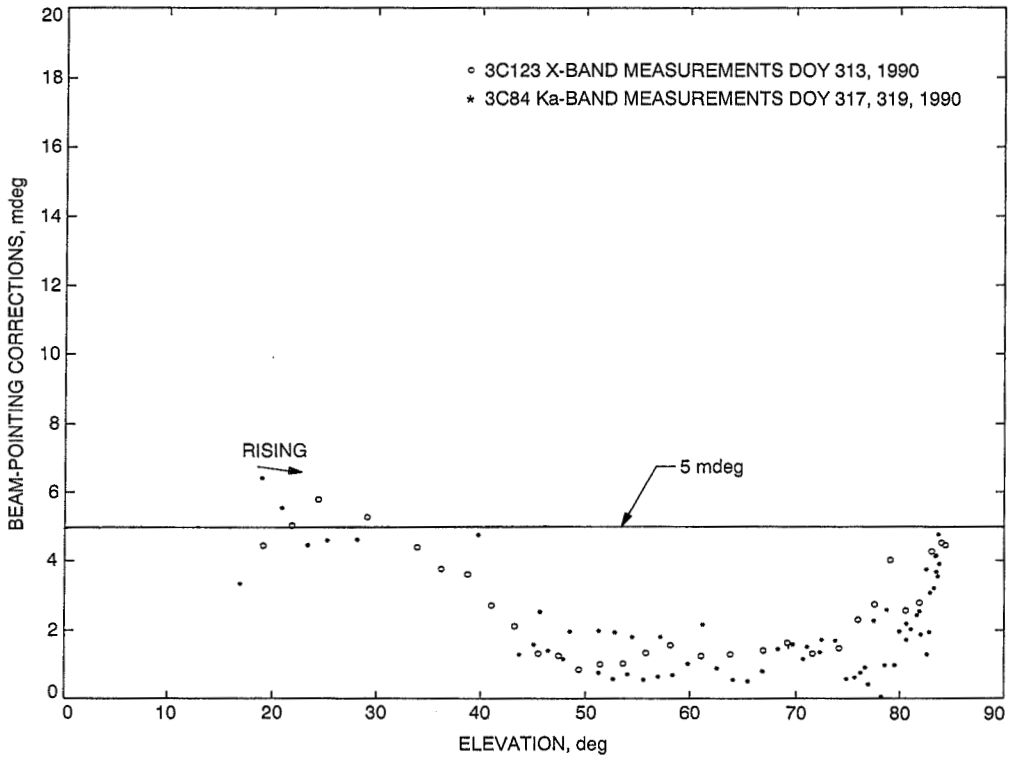


Fig. 15. Beam-pointing corrections for 3C123 and 3C84 tracks at f3.

Modeling and Simulations of the DSS 13 Antenna Control System

W. Gawronski and J. A. Mellstrom
Ground Antennas and Facilities Engineering Section

A model of the antenna control system for the azimuth and elevation axes of the DSS 13 antenna is developed. This model is used for simulation of elevation and azimuth dynamics, cross-coupled dynamics, and radio-frequency pointing error due to both input commands and wind disturbances. This model also serves as a tool for the antenna controller design. A modal state-space model of the antenna structure was obtained from its finite-element model with a free-rotating tipping structure and alidade. Model reduction techniques applied separately for the antenna structure, elevation and azimuth drives, and rate-loop model reduce the system order to one-third of that of the original, while preserving its significant dynamic properties. Extensive simulation results illustrate properties of the model.

I. Introduction

The control system model for the elevation drive of the DSS 13 antenna was described in [1]. This article extends the previous approach to the full control system of the antenna (in azimuth and elevation), and extends the structural model for modes up to 10 Hz. This article also describes the wind disturbance model used to study this important source of antenna pointing error. The antenna structural model for the 90-deg elevation angle is obtained from the antenna's finite-element model with a free rotation in azimuth and elevation. A linear structural model is analyzed. Nonlinear effects due to dry friction are not considered. A state-space model is used to describe the control system and its components.

The structure and the drives, as well as the rate-loop model, have some purely imaginary poles or poles at zero (the latter poles have integrating properties). Known reduction techniques fail when applied to systems with integrators. A balanced reduction technique, developed here for the systems with integrators, was applied to reduce the order of the structural model, elevation and azimuth drive models, and subsequently the rate-loop model. The order of the resulting reduced model is one-third of the original's, while the modeling error is small in comparison to the system dynamics.

The presented approach is an analytical base for computer software development. This software is designed as a tool for simulations of antenna dynamics due to the input commands and disturbances, for control system analysis, and for the position control algorithm design. Extensive simulations have been performed and described to illustrate the overall system properties.

II. Structural Model

The antenna structural model for elevation control purposes was described in [1]. In this article, structural modeling is expanded for the antenna in both elevation and azimuth motions. New developments specific to the joint azimuth and elevation modeling are presented; the reader is referred to [1] for the approaches used in the elevation control system model. The state-space representation denoted by a quadruple (A, B, C, D) is a set of first-order differential equations

$$\dot{x} = Ax + Bu, \qquad y = Cx + Du \tag{1}$$

where the state vector x is of dimension $n \times 1$, the input u is of dimension $p \times 1$, the output y is of dimension $q \times 1$, and the matrices A, B, C, and D are of dimensions $n \times n, n \times p, q \times n$, and $q \times p$, respectively.

A. Full Model

In this section the state-space quadruple in modal coordinates (A_s, B_s, C_s, D_s) for the antenna structure is determined. The state vector of the structure x_s consists of modal displacements q_s and modal velocities v_s,

$$x_s^T = [q_s^T \ v_s^T] \tag{2}$$

From the finite-element model, generated by the JPL–IDEAS code[1], $m = 21$ modes are obtained. The first and the second modes are rigid-body modes with zero natural frequency. The modes are determined for r selected points of interest; thus, $\phi_i = [\phi_{i1}, \phi_{i2}, \cdots, \phi_{ir}]^T, i = 1, \cdots, 21$. The antenna structural model is generated from the natural frequencies ω_i and mode shapes ϕ_i. Modal damping of the structure ζ_i is assumed to be $\zeta_i = 0.005$ for $i = 1, \cdots, 21$. The antenna structure is free to rotate about the elevation and azimuth axes.

1. Matrix A_S. Denote

$$\Omega = diag(\omega_i), \qquad Z = diag(\zeta_i), i = 1, 2, \cdots, m \tag{3}$$

then the system matrix A_s for the antenna structure is

$$A_s = \begin{bmatrix} 0 & I_m \\ -\Omega^2 & -2Z\Omega \end{bmatrix} \tag{4}$$

where I_m is the identity matrix of dimensions $m \times m$. Details of the derivation of A_s are presented in [1]. Numerical values of Ω, Z, and other parameters are given in Appendix A.

2. Matrix C_S. The following outputs of the antenna structure are considered: elevation angle θ_e, elevation rate $\dot{\theta}_e$, elevation pinion rate $\dot{\theta}_{pe}$, azimuth angle θ_a, azimuth rate $\dot{\theta}_a$, azimuth pinion rates $\dot{\theta}_{pa1}$ and $\dot{\theta}_{pa2}$, pointing error angle in elevation ϵ_{el}, pointing error angle in cross-elevation ϵ_{xel}, and subreflector x, y, and z positions q_{sx}, q_{sy}, and q_{sz}. Thus, the antenna output matrix C_s consists of twelve rows:

$$C_s^T = [C_{s1}^T, C_{s2}^T, \cdots, C_{s11}^T, C_{s12}^T] \tag{5}$$

where

C_{s1} is the elevation encoder reading

C_{s2} is the elevation rate

C_{s3} is the elevation pinion rate

C_{s4} is the azimuth encoder reading

C_{s5} is the azimuth rate

C_{s6} and C_{s7} are azimuth pinion rates

C_{s8} and C_{s9} are the pointing errors in elevation and cross-elevation

C_{s10}, C_{s11}, and C_{s12} are the subreflector x, y, and z positions

The first two rows (the elevation angle and rate) are determined as follows. In the finite-element model, the node at the bull gear center has a label $nb = 5380$ and the node at distance R to nb has a label $nc = 41212$ (see Fig. 1). The high stiffness of the bull gear and the close location of the two nodes allows one to determine the elevation angle as a rigid-body rotation

$$\theta_e = \frac{y_{nc} - y_{nb}}{R} = \frac{(C_{ncy}\Phi - C_{nby}\Phi)q_s}{R} = \frac{(\phi_{ncy} - \phi_{nby})q_s}{R}$$

where R is the bull gear radius and C_{ncy} and C_{nby} denote row vectors with one nonzero element. The nonzero element of C_{ncy} is equal to 1 and is located in the position in q_s corresponding to the y displacement of node nc. The

[1] R. Levy and D. Strain, *JPL–IDEAS Finite Element Analysis and Design Optimization Program*, Document NPO-17783 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1988.

nonzero element of C_{nby} is equal to 1 and is located in the position corresponding to the y displacement at node nb in the vector q_s. Thus, ϕ_{ncy} and ϕ_{nby} are

$$\phi_{ncy} = [\phi_{1ncy}, \cdots, \phi_{mncy}], \qquad \phi_{nby} = [\phi_{1nby}, \cdots, \phi_{mnby}]$$

where ϕ_{incy} is the ith mode component at node nc in the y direction and ϕ_{inby} is the ith mode component at node nb in the y direction. From this, it follows that:

$$C_{s1} = [0\ C_{se}], \qquad C_{s2} = [C_{se}\ 0] \tag{6a}$$

where $C_{se} = (\phi_{ncy} - \phi_{nby})/R$.

For the determination of the pinion rate measurement matrix C_{s3}, denote the velocity at pinion housing v_p (at node nu), bull gear velocity v_b at node no (see Fig. 1), and their projections v_1 and v_2 onto a plane tangential to the bull gear at node no. For the pinion rate $\dot{\theta}_{pe} = (v_1 - v_2)/r_{pe}$ and for $v_1 = C_{v1}v_s$ and $v_2 = C_{v2}v_s$ from the JPL–IDEAS code, one obtains

$$v_1 - v_2 = \frac{C_{soe}v_s}{\dot{r}_{pe}}$$

where v_s is a modal velocity defined in Eq. (2) and $C_{soe} = C_{v1} - C_{v2}$. Therefore, $\dot{\theta}_{pe} = C_{spe}v_s = [0\ C_{spe}]x_s$, where $C_{spe} = C_{soe}/r_{pe}$, finally giving

$$C_{s3} = [0\ C_{spe}] \tag{6b}$$

The matrix C_{soe} is directly obtained from the JPL–IDEAS code.

The azimuth position angle is obtained as $\theta_a = C_a q_s$, where C_a is a result of the JPL–IDEAS code and, thus, $\theta_a = [C_a\ 0]x_s, \dot{\theta}_a = [0\ C_a]x_s$. Therefore,

$$C_{s4} = [C_a\ 0], \qquad C_{s5} = [0\ C_a] \tag{6c}$$

The last seven rows of C_s are obtained from the finite-element model. For azimuth pinion rates $\dot{\theta}_{pa1} = [0\ C_{pa1}]x_s, \dot{\theta}_{pa2} = [0\ C_{pa2}]x_s$, pointing errors $\epsilon_{el} = [C_{ere}\ 0]$, $\epsilon_{xel} = [C_{erx}\ 0]$, and subreflector positions $q_{sx} = [C_{sx}\ 0]$, $q_{sy} = [C_{sy}\ 0]$, and $q_{sz} = [C_{sz}\ 0]$, thus,

$$C_{s6} = [0\ C_{pa1}], \qquad C_{s7} = [0\ C_{pa2}] \tag{6d}$$

$$C_{s8} = [C_{ere}\ 0] \qquad C_{s9} = [C_{erx}\ 0] \tag{6e}$$

$$C_{s10} = [C_{sx}\ 0], \qquad C_{s11} = [C_{sy}\ 0], \qquad C_{s12} = [C_{sz}\ 0] \tag{6f}$$

3. Matrix B_S. The inputs to the structure are: an elevation-axis drive torque, two azimuth-axis drive torques, and a wind disturbance force. The elevation-axis drive torque is applied to the elevation bull gear by the pinion, as shown in Fig. 1. A node labeled $no = 86302$ in the finite-element model is the point of contact between the pinion and the bull gear, while a node with the label $nu = 86881$ is located at the joint between the supporting truss structure and the pinion housing. The rigid pinion housing includes two drive systems. Consider forces at nodes no and nu and their components tangential to the bull gear at node no (F_{tno}) and at node nu (F_{tnu}). Assuming a rigid pinion, the torque applied to the pinion of the gearbox is

$$T = \frac{F_{tno} - F_{tnu}}{r_{pe}}$$

where r_{pe} is the elevation pinion radius. From the JPL–IDEAS code, one obtains the relative tangential displacement Δq_{te} of nodes no and nu, in modal coordinates,

$$\Delta q_{te} = q_{tno} - q_{tnu} = C_{soe}q_s$$

where the matrix C_{soe} has been already described. Thus, the matrix B_{soe} for the force $\Delta F_{te} = F_{tno} - F_{tnu}$ is

$$B_{soe} = \frac{M_m^{-1}C_{soe}^T}{r_{pe}}$$

and the matrix B_{se} for the elevation torque is

$$B_{se} = \begin{bmatrix} 0 \\ \\ B_{soe} \end{bmatrix} \tag{7a}$$

The matrices B_{sa1} and B_{sa2} for the azimuth-axis drive torques are obtained as follows. The tangential displacements of the azimuth pinion contact points are given by

$$q_{t1} = C_{sa1}q_s, \qquad q_{t2} = C_{sa2}q_s$$

where q_{t1} and q_{t2} are the tangential displacements, C_{sa1} and C_{sa2} are determined from the JPL–IDEAS finite-element model, and q_s is the modal displacement vector of the structure. Thus, the azimuth-axis drive input matrices are

$$B_{sa1} = \begin{bmatrix} 0 \\ M_m^{-1} C_{sa1}^T \end{bmatrix}, \qquad B_{sa2} = \begin{bmatrix} 0 \\ M_m^{-1} C_{sa2}^T \end{bmatrix} \tag{7b}$$

Wind disturbances are applied as wind forces from the x and y directions (x is the elevation axis direction and y is a horizontal direction orthogonal to the elevation axis). The input matrices B_{wx} and B_{wy} for the wind forces are obtained from the JPL–IDEAS model; hence, the wind input matrix is

$$B_{wind} = \begin{bmatrix} 0 & 0 \\ B_{wx} & B_{wy} \end{bmatrix} \tag{7c}$$

4. Matrix D_S. Matrix D_s is a zero matrix of dimensions $q \times p$, where q is the number of outputs and p is the number of inputs to the structure.

B. Model Reduction

The structural model under consideration consists of 21 modes or 42 states. However, some of these modes are insignificant and can be eliminated. Observability and controllability properties are used to determine modes for elimination. A balanced representation [2] is a state-space representation with its states equally controllable and observable, and the Hankel singular value is the measure of the joint controllability and observability of each balanced state variable. The states with small Hankel singular values can be deleted, since they are weakly excited and weakly observed at the same time, causing a small modeling error. For flexible structures with small damping and distinct poles, the modal representation is almost balanced, cf. [3,4,5]. Each mode has almost the same controllability and observability property; hence, each mode can be considered for reduction separately. For a structure with m modes, matrix B_s has $2m$ rows and matrix C_s has $2m$ columns. Denote b_s as the last m rows of B_s, c_q as the first m columns of C_s, and c_r as the last m columns of C_s, then b_{si} is the ith row of b_s, c_{qi} is the ith column of c_q, and c_{ri} is the ith column of c_r. The Hankel singular value for the ith mode is given by Eq. (53) of [4] and Eq. (14) of [5]

$$\gamma_i = \frac{\sqrt{(w_{bi} b_{si} b_{si}^T)(w_{qi} c_{qi}^T c_{qi} + w_{ri} \omega_i^2 c_{ri}^T c_{ri})}}{4 \zeta_i \omega_i^2} \tag{8}$$

Care should be taken when determining Hankel singular values. Units should be consistent, otherwise some inputs or outputs receive more weight in Hankel singular value determination than necessary. Consider, for example, the azimuth encoder reading in arcseconds and the elevation encoder reading in degrees. For the same angle, the numerical reading of the azimuth encoder is 3600 larger than the elevation encoder reading; hence, the elements for the azimuth output are much larger than those for elevation. On the other hand, some variables need more attention than others: the pointing error and encoder readings are the most important factors in the antenna performance; hence, their importance has to be emphasized in mode evaluation. For these two reasons, consistency of units and variable importance, the weighting factors $w_{bi} > 0, w_{qi} > 0$, and $w_{ri} > 0$ for $i = 1, \cdots, m$ are introduced. Typically, weights are set to 1. However, for more important variables, the weight is set larger than 1.

For each mode, the Hankel singular value is determined and used to decide the number of modes in the reduced structural model. For the rigid-body modes, Hankel singular values tend to infinity; hence, rigid-body modes are always included in the reduced model. Hankel singular values, with all weights set to one, of the 19 flexible modes of the antenna model are plotted in Fig. 2.

The reduced order model consists of 10 modes: two rigid-body modes and eight elastic modes, with natural frequencies 3.1240, 3.4880, 4.1140, 4.4450, 6.5820, 7.1540, 7.4880, and 9.112 Hz (shadowed in Fig. 2). The transfer functions of the full and reduced models of the antenna are shown in Fig. 3, indicating that the reduced model represents properties of the original system. At this stage, wind disturbances are not included in the model and, therefore, are not considered in model reduction. One should note, however, that for a model with disturbances included, a different reduced model will be obtained, as shown in Section VI of this article.

III. Drive System Model

Three antenna drives, the elevation drive (ELD) and two azimuth drives (AZD1 and AZD2), are analyzed. Each of the drives has three inputs: rate command ($\dot{\theta}_{oe}$ or $\dot{\theta}_{oa}$), torque bias (v_{bse} or v_{bsa1}, v_{bsa2}), and pinion rate ($\dot{\theta}_{pe}$ or $\dot{\theta}_{pa1}, \dot{\theta}_{pa2}$), and the following outputs: torques (T_e or

T_{a1}, T_{a2}) and bias voltages (v_{c1}, v_{c2}) to balance the load between two azimuth drives. The amplification parameter k_a is given in Appendix A.

A. Full Model

The structure of each drive is the same, as shown in Fig. 4. The structure consists of two identical subsystems denoted G_o (which model the motor, amplifier, gearbox, tachometer, and rate-loop compensation electronics), the command amplifier G_c, and the torque share bias loop. The subsystem G_c is shown in Fig. 5. Denoting the first state variable $x_1 = v_1$, the G_c subsystem's state equation is

$$\dot{x}_1 = -\tau_1^{-1} x_1 + k_1 k_s \tau_1^{-1} \dot{\theta}_o \tag{9}$$

The equations for subsystem G_o are derived in [1]. The inputs are a biased rate command (v_1) and pinion rate ($\dot{\theta}_p$). An additional brake parameter κ, will be discussed later. The outputs are motor current (i_o), motor angular velocity (ω_m), and torque (T). These outputs are the fifth, sixth, and seventh states, respectively, in the following state-space equations for subsystem G_o

$$\dot{x}_o = A_o x_o + B_{11} v_1 + B_{12} \dot{\theta}, \qquad \begin{bmatrix} T \\ i_o \end{bmatrix} = C_o x_o \tag{10}$$

where

$$A_o = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{\tau_3} & 0 & 0 & 0 & -k_{tach} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{k_r}{\tau_3} & \frac{k_r \tau_2}{\tau_3} & 0 & -\frac{1}{\tau_5} & -k_{cur} & 0 & 0 \\ 0 & 0 & \frac{k_f k_i}{L_a \tau_5} & \frac{k_f k_i \tau_4}{L_a \tau_5} & -\frac{R_a}{L_a} & -\frac{k_b}{L_a} & 0 \\ 0 & 0 & 0 & 0 & \frac{k_m}{J_m} & 0 & -\frac{1}{J_m N} \\ 0 & 0 & 0 & 0 & 0 & k_{gx} N_x & 0 \end{bmatrix}$$

$$B_{11}^T = [0\ 1\ 0\ 0\ 0\ 0\ 0],\ B_{12}^T = [0\ 0\ 0\ 0\ 0\ 0\ -k_{gx} N_x N]$$

$$C_{11} = [0\ 0\ 0\ 0\ 0\ 0\ 1],\ C_{12} = [0\ 0\ 0\ 0\ 1\ 0\ 0]$$

$$C_o = \begin{bmatrix} C_{11} \\ C_{12} \end{bmatrix}, \qquad B_o = [B_{11}\ B_{12}]$$

Two systems' G_o's are located in the drive system, see Fig. 4: the first one with inputs $v_{11}, \dot{\theta}_p$, outputs T_1, i_{o1}, and state vector x_{21}; and the second one with inputs $v_{12}, \dot{\theta}_p$, outputs T_2, i_{o2}, and state vector x_{22}. The equations for these systems are

$$\dot{x}_{21} = A_o x_{21} + B_{11} v_{11} + B_{12} \dot{\theta}, \qquad \begin{bmatrix} T_1 \\ i_{o1} \end{bmatrix} = C_o x_{21} \tag{11}$$

$$\dot{x}_{22} = A_o x_{22} + B_{11} v_{12} + B_{12} \dot{\theta}, \qquad \begin{bmatrix} T_2 \\ i_{o2} \end{bmatrix} = C_o x_{22} \tag{12}$$

The amplifier for the bias signal has an input of Δv_c and an output of v_u. Denoting $x_3 = v_u$, this amplifier's state-space equation is obtained

$$\dot{x}_3 = -\tau_6^{-1} x_3 + k_{ctfr} \tau_6^{-1} \Delta v_c \tag{13}$$

The nodal equations complete the set of the drive-system model equations

$$T = T_1 + T_2, \quad T_1 = C_{11} x_{21}, \quad T_2 = C_{11} x_{22} \tag{14a}$$

$$\begin{aligned} \Delta v_c &= v_{c1} - v_{c2} + k_{bs} v_{bias} \\ &= k_{ct} C_{12} (x_{21} - x_{22}) + k_{bs} v_{bias} \end{aligned} \tag{14b}$$

$$v_{11} = v_1 - v_u = x_1 - x_3, \qquad v_{12} = v_1 + v_u = x_1 + x_3 \tag{14c}$$

$$v_c = v_{c1} + v_{c2} = k_{ct} C_{12} (x_{12} + x_{22}) \tag{14d}$$

Denoting the state vector $x_m^T = [x_1^T, x_{21}^T, x_{22}^T, x_3^T]$ and combining Eqs. (9) through (14), one can write the equations for the drive system as follows:

$$\dot{x}_m = A_m x_m + B_m u_m, \qquad y_m = C_m x_m \tag{15}$$

where $u_m^T = [\dot{\theta}_o\ \dot{\theta}_p\ v_{bias}]$, $y_m^T = [T\ v_c\ \omega_{m1}\ \omega_{m2}]$, and

$$A_m = \begin{bmatrix} -1/\tau_1 & 0 & 0 & 0 \\ B_{11} & A_o & 0 & -B_{11} \\ B_{11} & 0 & A_o & B_{11} \\ 0 & p_1 & -p_1 & -1/\tau_6 \end{bmatrix}$$

$$B_m = [B_{mo}\ B_{mp}\ B_{mbs}] = \begin{bmatrix} k_1 k_s/\tau_1 & 0 & 0 \\ 0 & B_{12} & 0 \\ 0 & B_{12} & 0 \\ 0 & 0 & p_2 \end{bmatrix}$$

$$C_m = \begin{bmatrix} C_{mt} \\ C_{mc} \\ C_{om1} \\ C_{om2} \end{bmatrix} = \begin{bmatrix} 0 & C_{11} & C_{11} & 0 \\ 0 & k_{ct}C_{12} & k_{ct}C_{12} & 0 \\ 0 & C_{13} & 0 & 0 \\ 0 & 0 & C_{13} & 0 \end{bmatrix}$$

$$p_1 = \frac{k_{ctfr}k_{ct}C_{12}}{\tau_6}, \qquad p_2 = \frac{k_{ctfr}k_{ct}}{\tau_6}$$

$$C_{13} = [0\ 0\ 0\ 0\ 0\ 1\ 0]$$

B. Model Reduction

The drive system model is reduced by applying the balancing principle. The system has a pole at zero; hence, the grammians and Hankel singular values do not exist, but tend toward infinity. It is shown in Appendix B that the balanced representation exists for the case of poles at zero. The reduction using antigrammians can be completed. Note that the reduction of a triple (A, B, C) leads now to the quadruple (A_r, B_r, C_r, D_r). This fact has to be taken into account when deriving the rate-loop model.

The drive system has 16 state variables. The plots of the singular values of the balaneed antigrammians for the elevation and azimuth drive systems are shown in Fig. 6. The singular values of component number 5 and greater are large in comparison to the remaining components. Therefore, the reduced system consists of five state variables (one state with the pole at zero). The transfer functions of the reduced and full elevation drive systems are compared in Fig. 7 (transfer function for the azimuth drives are similar). From Fig. 7, one can see that the reduced model almost exactly approximates the full model.

IV. Rate-Loop Model

In this section a rate-loop model with active azimuth and elevation drives is presented. Also, a model with an active elevation drive and with activated azimuth brakes is derived, along with a model with an active azimuth drive and with elevation brakes activated. The obtained rate-loop models are reduced to obtain the smallest acceptable model.

A. Full Model

The rate-loop model of the antenna is shown in Fig. 8. The model consists of the antenna structure model and models of three drives: the elevation drive (ELD) and two azimuth drives (AZD1 and AZD2) described above.

Combining the equations for structure and the ELD, AZD1, and AZD2 with the nodal equations for the rate-loop model as in Fig. 10, one obtains the following equations (where the term D appeared as a result of model reduction). For the elevation drive,

$$\dot{x}_e = A_{me}x_e + [B_{moe}\ B_{mpe}\ B_{mbse}] \begin{bmatrix} \dot{\theta}_{oe} \\ \dot{\theta}_{pe} \\ v_{bse} \end{bmatrix} \tag{16a}$$

$$T_e = C_{mt}x_e + [D_{eo}\ D_{ep}\ D_{ebs}] \begin{bmatrix} \dot{\theta}_{oe} \\ \dot{\theta}_{pe} \\ v_{bse} \end{bmatrix} \tag{16b}$$

For azimuth drive AZD1,

$$\dot{x}_{a1} = A_{ma}x_{a1} + [B_{moa}\ B_{mpa1}\ B_{mbsa}] \begin{bmatrix} \dot{\theta}_{a1} \\ \dot{\theta}_{pa1} \\ v_{bsa1} \end{bmatrix} \tag{17a}$$

$$\begin{bmatrix} T_{a1} \\ v_{c1} \end{bmatrix} = \begin{bmatrix} C_{mt} \\ C_{mc} \end{bmatrix} x_{a1} + \begin{bmatrix} D_{aot} & D_{apt} & D_{abst} \\ D_{aov} & D_{apv} & D_{absv} \end{bmatrix} \begin{bmatrix} \dot{\theta}_{a1} \\ \dot{\theta}_{pa1} \\ v_{bsa1} \end{bmatrix} \tag{17b}$$

For azimuth drive AZD2,

$$\dot{x}_{a2} = A_{ma}x_{a2} + [B_{moa}\ B_{mpa2}\ B_{mbsa}]\begin{bmatrix}\dot{\theta}_{a2}\\ \dot{\theta}_{pa2}\\ v_{bsa2}\end{bmatrix} \quad (18a)$$

$$\begin{bmatrix}T_{a2}\\ v_{c2}\end{bmatrix} = \begin{bmatrix}C_{mt}\\ C_{mc}\end{bmatrix} x_{a2} + \begin{bmatrix}D_{aot} & D_{apt} & D_{abst}\\ D_{aov} & D_{apv} & D_{absv}\end{bmatrix}\begin{bmatrix}\dot{\theta}_{a2}\\ \dot{\theta}_{pa2}\\ v_{bsa2}\end{bmatrix} \quad (18b)$$

For the antenna structure,

$$y_s = C_s x_s \quad (19a)$$

$$\dot{x}_s = A_s x_s + B_{se}T_e + B_{sa1}T_{a1} + B_{sa2}T_{a2} + B_{wind}f_{wind} \quad (19b)$$

The nodal equations are

$$\dot{\theta}_{a1} = \dot{\theta}_{oa} - v_{aa} = \dot{\theta}_{oa} - k_a(v_{c1} - v_{c2}) = \dot{\theta}_{oa} - k_a C_{mc}(x_{a1} - x_{a2}) \quad (20a)$$

$$\dot{\theta}_{a2} = \dot{\theta}_{oa} + v_{aa} = \dot{\theta}_{oa} + k_a(v_{c1} - v_{c2}) = \dot{\theta}_{oa} + k_a C_{mc}(x_{a1} - x_{a2}) \quad (20b)$$

For the state vector $x_{rl}^T = [x_e^T\ x_{a1}^T\ x_{a2}^T\ x_s^T]$, the input $u_{rl} = [\dot{\theta}_{oe}\ \dot{\theta}_{oa}\ v_{bse}\ v_{bsa1}\ v_{bsa2}\ f_{wind}]^T$, the output $y_{rl} = [\theta_e\ \theta_a]^T$, and the state triple (A_{rl}, B_{rl}, C_{rl}) is found

$$A_{rl} = \begin{bmatrix} A_{me} & 0 & 0 & B_{mpe}C_{s3}\\ 0 & A_{ma} - k_a B_{moa}C_{mca} & k_a B_{moa}C_{mca} & B_{mpa}C_{s6}\\ 0 & k_a B_{moa}C_{mca} & A_{ma} - k_a B_{moa}C_{mca} & B_{mpa}C_{s7}\\ B_{se}C_{mte} & B_{sa1}C_{mta} + k_a \times (B_{sa2} - B_{sa1})D_{aot}C_{mca} & B_{sa2}C_{mta} + k_a \times (B_{sa1} - B_{sa2})D_{aot}C_{mca} & A_s + B_{se}D_{ep}C_{s3} + B_{sa1}D_{apt}C_{s6} + B_{sa2}D_{apt}C_{s7} \end{bmatrix} \quad (21a)$$

$$B_{rl} = \begin{bmatrix} B_{moe} & 0 & B_{mbse} & 0 & 0 & 0\\ 0 & B_{moa} & 0 & B_{mbsa} & 0 & 0\\ 0 & B_{moa} & 0 & 0 & B_{mbsa} & 0\\ B_{se}D_{eo} & (B_{sa1} + B_{sa2})D_{aot} & B_{se}D_{ebs} & B_{sa1}D_{abst} & B_{sa2}D_{abst} & B_{wind} \end{bmatrix} \quad (21b)$$

$$C_{rl} = \begin{bmatrix} 0 & 0 & 0 & C_{s1}\\ 0 & 0 & 0 & C_{s4} \end{bmatrix} \quad (21c)$$

B. Rate-Loop Model With Active Elevation or Azimuth Brakes

Most frequently, the antenna dynamics are simulated with elevation and azimuth drives active simultaneously.

There are situations, however, when either the azimuth or elevation brakes are on and only one antenna drive is active. There are two ways of modeling the brake action in this case. In the first way, the numerical value to the parameter κ, shown in Fig. 5, is set. For an active brake, the tachometer shaft is still, hence $\omega_m = 0$. This is obtained by setting $\kappa = 0$ in Fig. 5. For the case where the brake is off, one sets $\kappa = 1$. Despite its simplicity, this approach leads to an unnecessarily large model. The dynamics of the drive with the active brakes are included in the model, although the drive is not active. This disadvantage is removed by deriving separate models for the antenna with active brakes.

Block diagrams of the antenna control systems with active brakes are shown in Fig. 9. For the active elevation brakes, the elevation torque T_e is due to elastic deformation of the pinion shaft, hence,

$$T_e = k_{goe}\theta_{pe} = k_{goe}[C_{spe}\ 0]x_s$$

Thus, the equation for the antenna structure is

$$\dot{x}_s = A_s x_s + \begin{bmatrix} 0 \\ B_{soe} \end{bmatrix} [C_{spe}\ 0]x_s + B_{sa1}T_{a1} + B_{sa2}T_{a2} \quad (22a)$$

and the following matrix for active elevation brakes is obtained:

$$A_{se} = A_s + \begin{bmatrix} 0 & 0 \\ B_{soe}C_{spe} & 0 \end{bmatrix} \quad (22b)$$

For the case of active elevation brakes, the state variable x_e representing the elevation drive is zero. These states are no longer necessary, and may be removed from the state vector x_{rl}. As a result: the first row and column of matrix A_{rl} are deleted and the matrix A_s is replaced with A_{se}; the first row and the first and third columns of matrix B_{rl} are deleted (no T_e and v_{bse} inputs); and the first column of matrix C_{rl} is removed. The block diagram of the rate-loop system with active elevation brakes is shown in Fig. 9(a).

As in the case for elevation, the torques at the azimuth drive, due to active azimuth brakes, are determined through the azimuth drive's stiffness and pinion angle of rotation

$$T_{a1} = k_{goa}\theta_{pa1} = k_{goa}[C_{spa1}\ 0]x_s$$

$$T_{a2} = k_{goa}\theta_{pa2} = k_{goa}[C_{spa2}\ 0]x_s$$

The resulting structure equation is

$$\dot{x}_s = A_s x_s + \begin{bmatrix} 0 \\ B_1 \end{bmatrix} [C_{spa1}\ 0]x_s + \begin{bmatrix} 0 \\ B_2 \end{bmatrix} [C_{spa2}\ 0]x_s + B_{se}T_e \quad (23a)$$

Thus, the structure matrix with azimuth brakes active is

$$A_{sa} = A_s + \begin{bmatrix} 0 & 0 \\ B_1C_{spa1} + B_2C_{spa2} & 0 \end{bmatrix} \quad (23b)$$

For the case where the azimuth brakes are active, the state variables x_{a1} and x_{a2}, representing the states of the azimuth drives, are zero. These states may be removed from the state vector x_{rl}. The rate-loop matrix A_{rl} in Eq. (21a) in this case has the second and the third rows and columns deleted and the matrix A_s is replaced with A_{sa}; the matrix B_{rl} has its second and third rows, and second, fourth, and fifth columns deleted (no T_{a1}, T_{a2}, and v_{bsa} inputs); and matrix C_{rl} has its second and third column deleted. The block diagram of the rate-loop system with active elevation brakes is shown in Fig. 9(b).

C. Model Reduction

Some of the poles of the rate-loop model are at zero; hence, the reduction procedure from Appendix B is applied. Note that the state triple (A_{rl}, B_{rl}, C_{rl}) is returned as a quadruple $(A_{rl}, B_{rl}, C_{rl}, D_{rl})$ after reduction

$$(A_{rl}, B_{rl}, C_{rl}) \Rightarrow (A_{rl}, B_{rl}, C_{rl}, D_{rl}) \quad (24)$$

Singular values of the balanced antigrammians of the rate-loop model are shown in Fig. 10. The rate-loop model designed from the reduced structural, elevation drive, and azimuth drive models consists of 35 states. It is reduced to 27 states—singular values of states 28 through 35 are large enough to have these states reduced. The frequency response of the reduced model is compared with the responses of the full model in Figs. 11 and 12 (the full model consists of a full antenna structural mode, and full azimuth and elevation drive models—all together 90 states).

V. Position Loop Model

The rate-loop system, with the position loop closed, is shown in Fig. 13. A proportional-plus-integral (PI) controller is applied. Consider first the series connection of the rate-loop system and the controller, as in Fig. 13(a). Define new state variables x_{ei} and x_{ai}

$$\dot{x}_{ei} = \theta_{eo} \text{ and } \dot{x}_{ai} = \theta_{ao} \tag{25}$$

and the state vector x_o for the series connection $x_o^T = [x_{rl}^T \; x_{ei} \; x_{ai}]$. The system output is

$$y_o = \begin{bmatrix} \theta_e \\ \theta_a \end{bmatrix} = \begin{bmatrix} C_e x_{rl} + d_{ee} r_e + d_{ea} r_a \\ C_a x_{rl} + d_{ae} r_e + d_{aa} r_a \end{bmatrix}$$

where d_{ee}, d_{ea}, d_{ae}, and d_{aa} are entries of D_{rl} obtained in Eq. (24). The input is $u_o^T = [\theta_{eo} \; \theta_{ao}]$ and the inputs to the rate-loop systems are

$$r_e = k_{ppe}\theta_{eo} + k_{iie}x_{ei} \text{ and } r_a = k_{ppa}\theta_{ao} + k_{iia}x_{ai} \tag{26}$$

where $k_{ppe}, k_{iie}, k_{ppa}$, and k_{iia} are proportional and integral parameters of the controllers. The equations for the rate-loop systems are

$$\dot{x}_{rl} = A_{rl}x_{rl} + B_e r_e + B_a r_a \tag{27}$$

Combining Eqs. (25), (26), and (27), one obtains

$$\dot{x}_o = A_o x_o + B_o u_o \text{ and } y_o = C_o x_o + D_o u_o \tag{28}$$

where

$$A_o = \begin{bmatrix} A_{rl} & k_{iie}B_e & k_{iia}B_a \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$B_o = \begin{bmatrix} k_{ppe}B_e & k_{ppa}B_a \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C_o = \begin{bmatrix} C_e & k_{iie}d_{ee} & k_{iia}d_{ea} \\ C_a & k_{iie}d_{ae} & k_{iia}d_{aa} \end{bmatrix}$$

$$D_o = \begin{bmatrix} k_{ppe}d_{ee} & k_{ppa}d_{ea} \\ k_{ppe}d_{ae} & k_{ppa}d_{aa} \end{bmatrix}$$

and B_e and B_a are the first and the second columns of B_{rl}.

For the closed-loop system

$$u_o = y_{com} - y_o \tag{29}$$

where $y_{com}^T = [\theta_{ecom} \; \theta_{acom}]$ is a command signal. Introducing Eq. (29) to Eq. (28), one obtains

$$\left.\begin{aligned} \dot{x}_{cl} &= A_{cl}x_{cl} + B_{cl}y_{com} + B_{clw}f_w \\ y_o &= C_{cl}x_{cl} + D_{cl}y_{com} \end{aligned}\right\} \tag{30}$$

The closed-loop system matrices are

$$\left.\begin{aligned} A_{cl} &= A_o - B_o C_o, \quad B_{cl} = B_o, \\ C_{cl} &= C_o - D_o C_o, \quad D_{cl} = D_o \end{aligned}\right\} \tag{31}$$

and the wind input matrix B_{clw} is obtained from the sixth (B_{r16}) and the seventh (B_{r17}) columns of B_{rl}

$$B_{clw} = \begin{bmatrix} B_{r16} & B_{r17} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The closed-loop system performance is illustrated in Figs. 14 and 15 in the frequency domain, and in Figs. 16 and 17 in the time domain.

VI. Wind Disturbance Simulations

In order to evaluate the antenna pointing error due to wind disturbances, two separate problems need to be investigated. First, wind forces acting on the antenna structure should be determined. Wind properties have been studied in [6,7], although wind forces acting on the antenna structure are not yet satisfactorily known. Secondly, the antenna model for simulation of the pointing error due to

wind disturbances should be developed. In this section, the latter problem is considered, presenting the wind disturbance simulation tools and simulation results for the assumed wind force properties.

The wind input to the antenna model is applied through the input matrix B_{wind} given in Section II. The entries of this matrix are large in comparison to the entries of the remaining input matrices B_{se}, B_{sa1}, and B_{sa2} for the elevation and azimuth pinion torques. The large values of B_{wind} appeared since the wind acts directly on the tipping structures, while torques act indirectly through the gearboxes, being reduced by the gearbox ratios. The large value of B_{wind} is a dominant factor in the reduction of the structural model; hence, it is compensated for by setting the appropriate weighting for the wind input—in this case the wind input weight is 10^{-8}. The input matrix B_{wind} is determined for a wind speed of 100 mph, and for this speed a unit wind force of $F_{ow} = 1$ lb is applied. The force F_w (lb) for the speed v_w (mph) is

$$F_w = \alpha_w^2 F_{ow} \text{ and } \alpha_w = \frac{v_w}{100} \tag{32}$$

The reduction procedure for the model with wind disturbances is similar to that presented above, however, the results are different. The plot of the Hankel singular values for the structure with the wind input is presented in Fig. 18. Comparing this plot with the one in Fig. 2 shows that a different set of modes plays the most important role in the system reduction. The wind input to the rate-loop model and the position loop model is given in Sections IV and V, respectively. The magnitude of the transfer function from the wind in the x and y directions to the elevation and cross-elevation pointing errors is shown in Fig. 19. The figures show that the flexible part of the antenna has a significant part in the pointing error balance.

Two kinds of antenna pointing error due to wind disturbances have been simulated: static error due to constant wind pressure, and dynamic error due to wind turbulence. The following results for the static wind load are obtained. For a 30-mph x-direction wind, the elevation pointing error is 2.1 mdeg and the cross-elevation pointing error is 8.9 mdeg. For a 30-mph y-direction wind, the elevation pointing error is 0.5 mdeg and the cross-elevation pointing error is 0.08 mdeg. The small pointing error due to the wind in the y direction is a result of the integrating action of the controller.

The dynamic wind force is generated as a random white-noise process pre-filtered by a filter with the spectral properties presented in Fig. 20. The obtained spectrum is an arbitrarily selected wind spectrum, a rough approximation of the Davenport model of the wind for a filter with $f_{max} = 0.1$ Hz, where f_{max} is the frequency of maximal value of the spectrum. Simulations have been performed for the position loop model consisting of 39 states. The forcing function is shown in Fig. 21, with a root mean square (rms) value equal to 0.018 lb (20 percent of the static pressure). The antenna pointing errors, in both elevation and cross-elevation, are presented in Fig. 22. For an x-direction wind, the rms error in elevation is 0.61 mdeg, and in cross-elevation it is 2.45 mdeg. For a y-direction wind, the rms error in elevation is 2.87 mdeg, while in cross-elevation, it is 0.21 mdeg. Note that the static pointing error is dominant for the x-direction wind, and the dynamic pointing error is dominant for the y-direction wind. The y-direction static action of the wind is sensed by the elevation encoder and is compensated for by the PI controller, while the static x-direction wind and the dynamic wind are barely registered by the encoder and are thus not compensated for by the controller. Again, the dynamic pointing error, due to some arbitrariness in selecting the filter, is an approximate result. More studies on wind force properties need to be done.

VII. Conclusions

In this article the modeling procedures and reduction techniques for the DSS 13 antenna structure and its control system have been presented. Through balancing the system controllability and observability properties, a reduced system model has been obtained. Antenna dynamics, rate-loop dynamics, and dynamics of a system with a closed position loop have been simulated. The model allows one to simulate elevation dynamics, cross-coupled dynamics in azimuth and elevation, and the dynamic radio-frequency pointing error due to command inputs and wind disturbances.

Acknowledgments

The authors would like to thank Robert Hill, Roy Levy, Ben Parvin, and Douglas Strain for their extensive and useful discussions.

References

[1] W. Gawronski and J. A. Mellstrom, "Elevation Control System Model for the DSS 13 Antenna," *TDA Progress Report 42-105*, vol. January–March 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 82–107, May 15, 1991.

[2] B. C. Moore, "Principal Component Analysis in Linear Systems: Controllability, Observability and Model Reduction," *IEEE Transactions Autom. Control*, vol. 26, no. 1, pp. 17–32, January 1981.

[3] E. A. Jonckheere, "Principal Component Analysis of Flexible Systems—Open Loop Case," *IEEE Transactions Autom. Control*, vol. 27, no. 12, pp. 1095–1097, December 1984.

[4] W. Gawronski and J.-N. Juang, "Model Reduction for Flexible Structures," in *Control and Dynamics Systems*, edited by C. T. Leondes, vol. 36, pp. 143–222, New York: Academic Press, 1990.

[5] W. Gawronski and T. Williams, "Model Reduction for Flexible Space Structures," *Journal of Guidance, Control, and Dynamics*, vol. 14, no. 1, pp. 68–76, January 1991.

[6] R. Levy and H. McGinnes, *Wind Power Prediction Models,* NASA Technical Memorandum 33-802, NASA, Washington, D.C., November 15, 1976.

[7] M. Berg, R. Levy, H. McGinnes, and D. Strain, *Wind Speed Statistics for Goldstone, California, Anemometer Sites*, JPL Publication 81-17, Jet Propulsion Laboratory, Pasadena, California, April 15, 1981.

[8] L. Fortuna, A. Gallo, and G. Nunnar, "New Results Involving Open-Loop Balanced Realization Schemes," *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, Florida, pp. 1331–1333, December 1989.

[9] S. Weiland, "Balanced Representations and Approximations of Linear Systems," *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, Florida, pp. 1334–1336, December 1989.

[10] C. S. Hsu and D. Hou, *Model Reduction of Unstable Linear Control Systems*, paper presented at the Third International Conference on Advances in Communication and Control Systems, Victoria, British Columbia, Canada, 1991.

[11] R. C. Dorf, *Modern Control Systems*, Reading, Massachusetts: Addison-Wesley, 1980.

[12] R. A. Horn and C. A. Johnson, *Matrix Analysis*, Cambridge, England: Cambridge University Press, 1985.

[13] Y. Liu and B. D. O. Anderson, "Singular Perturbation Approximation of Balanced Systems," *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, Florida, pp. 1355–1360, 1989.

[14] R. Prakash and S. Vittal Rao, "Model Reduction by Low Frequency Approximation of Internally Balanced Representation," *Proceedings of the 28th IEEE Conference on Decision and Control*, Tampa, Florida, pp. 2425–2430, 1989.

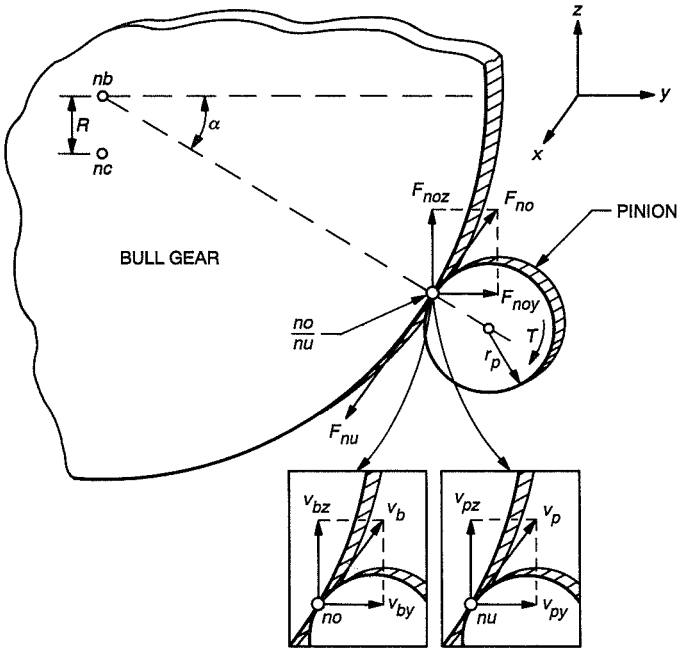


Fig. 1. Forces and velocities at the bull-gear–pinion connection.

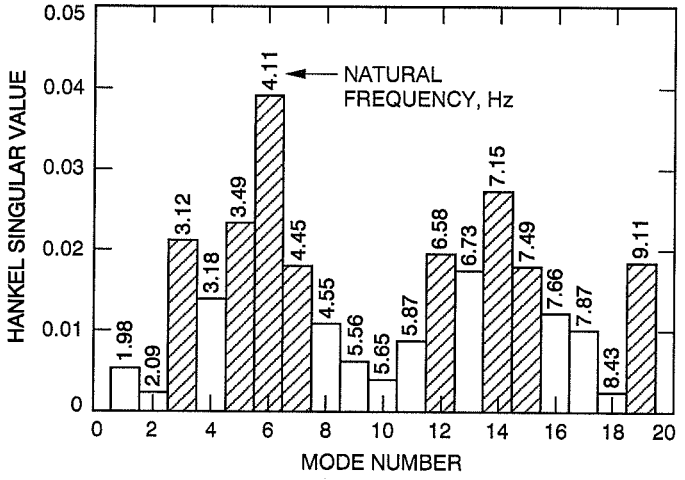


Fig. 2. Hankel singular values for the antenna structure.

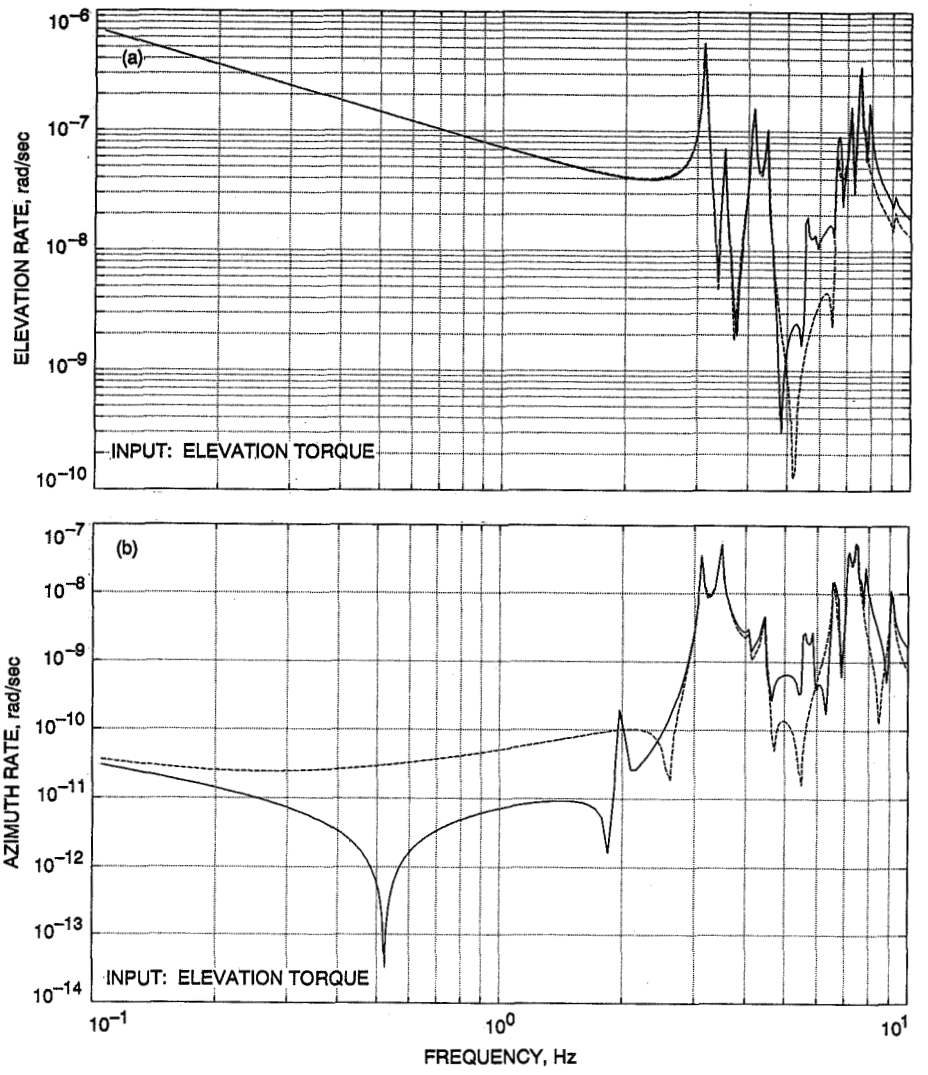


Fig. 3. Magnitudes of the transfer function of the full (42 states) (solid line) and the reduced (20 states) (dashed line) models of the antenna structure.

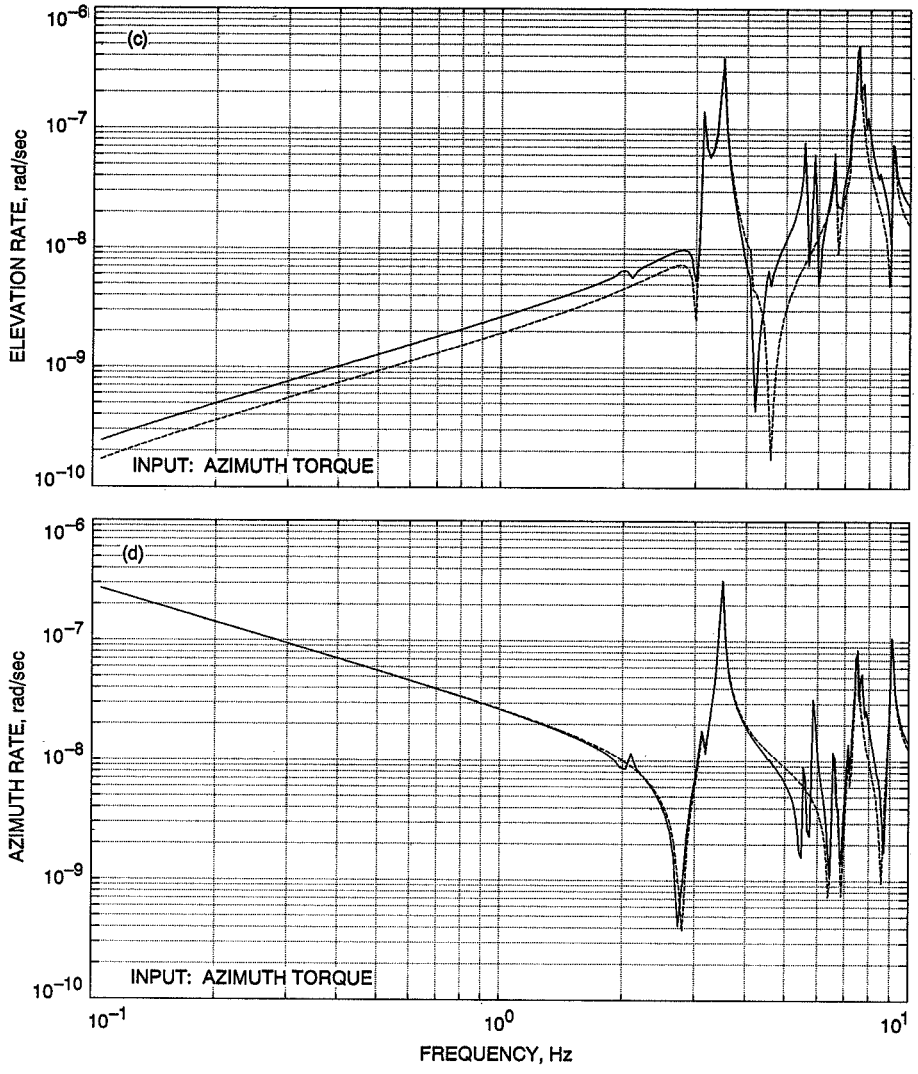

(c)
ELEVATION RATE, rad/sec
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
INPUT: AZIMUTH TORQUE
(d)
AZIMUTH RATE, rad/sec
10^{-6}
10^{-7}
10^{-8}
10^{-9}
10^{-10}
INPUT: AZIMUTH TORQUE
10^{-1}
10^{0}
10^{1}
FREQUENCY, Hz

Fig. 3 (contd).

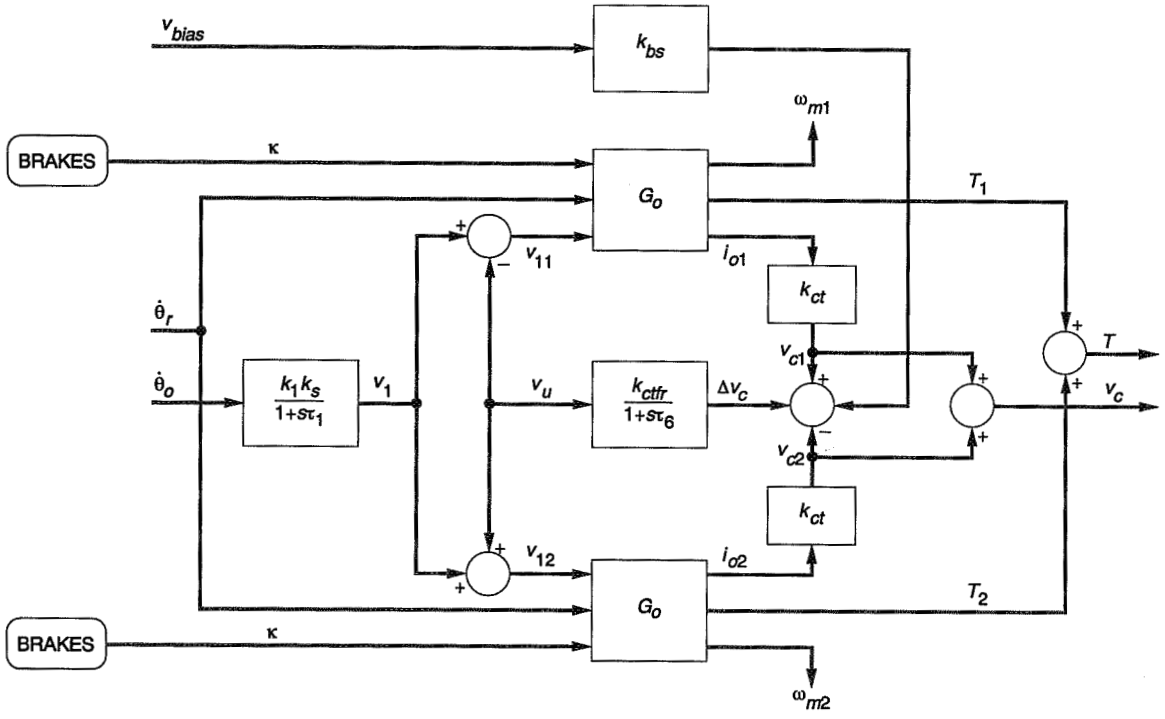


Fig. 4. Drive system model.

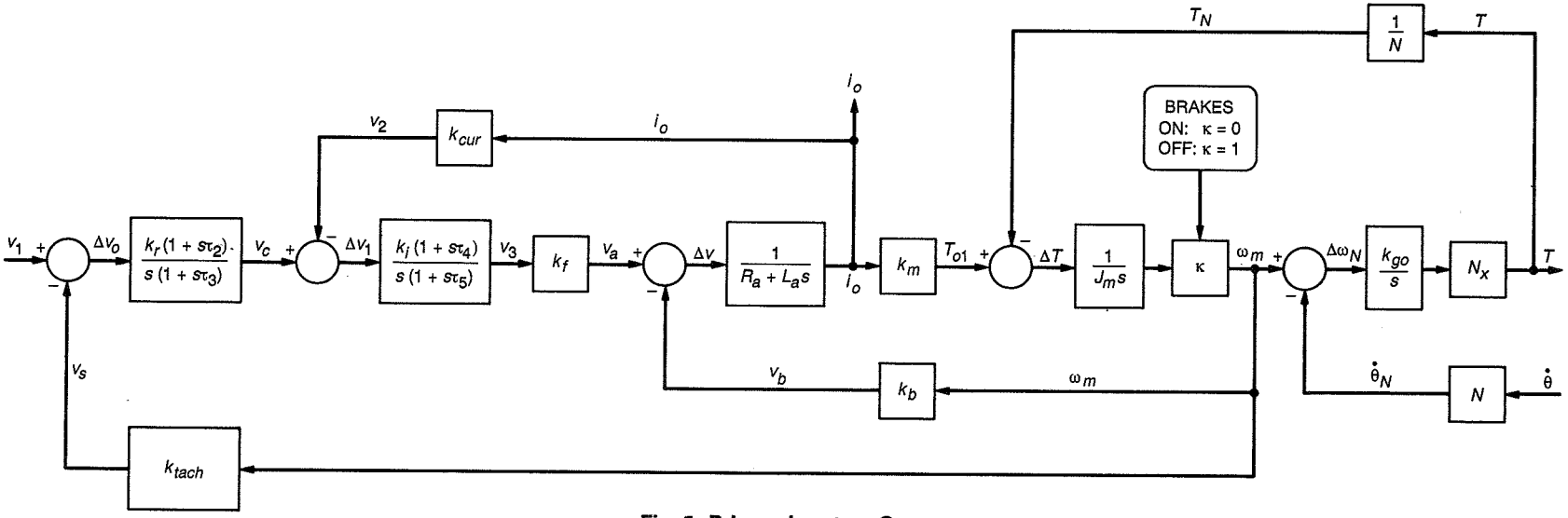


Fig. 5. Drive subsystem G_o.

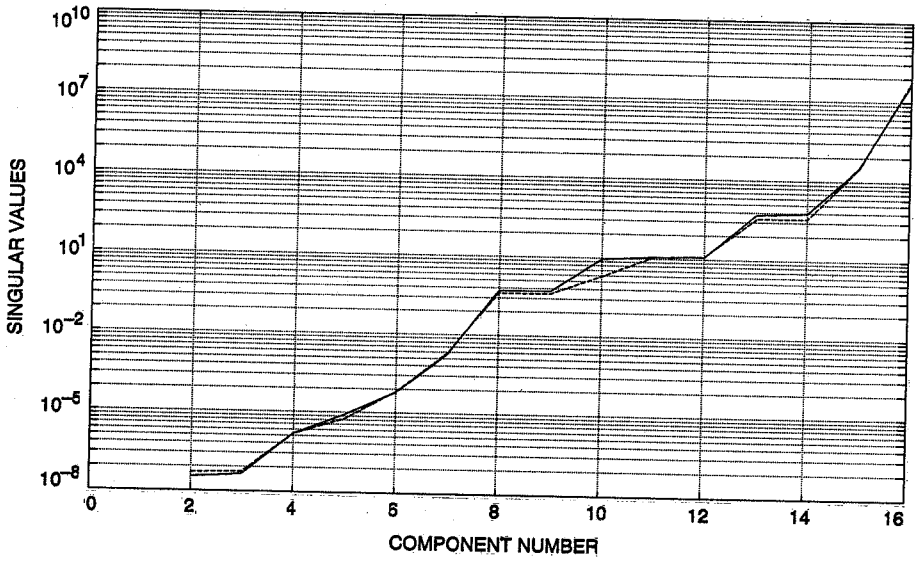


Fig. 6. Singular values of the balanced antigrammians for the elevation and azimuth drives.

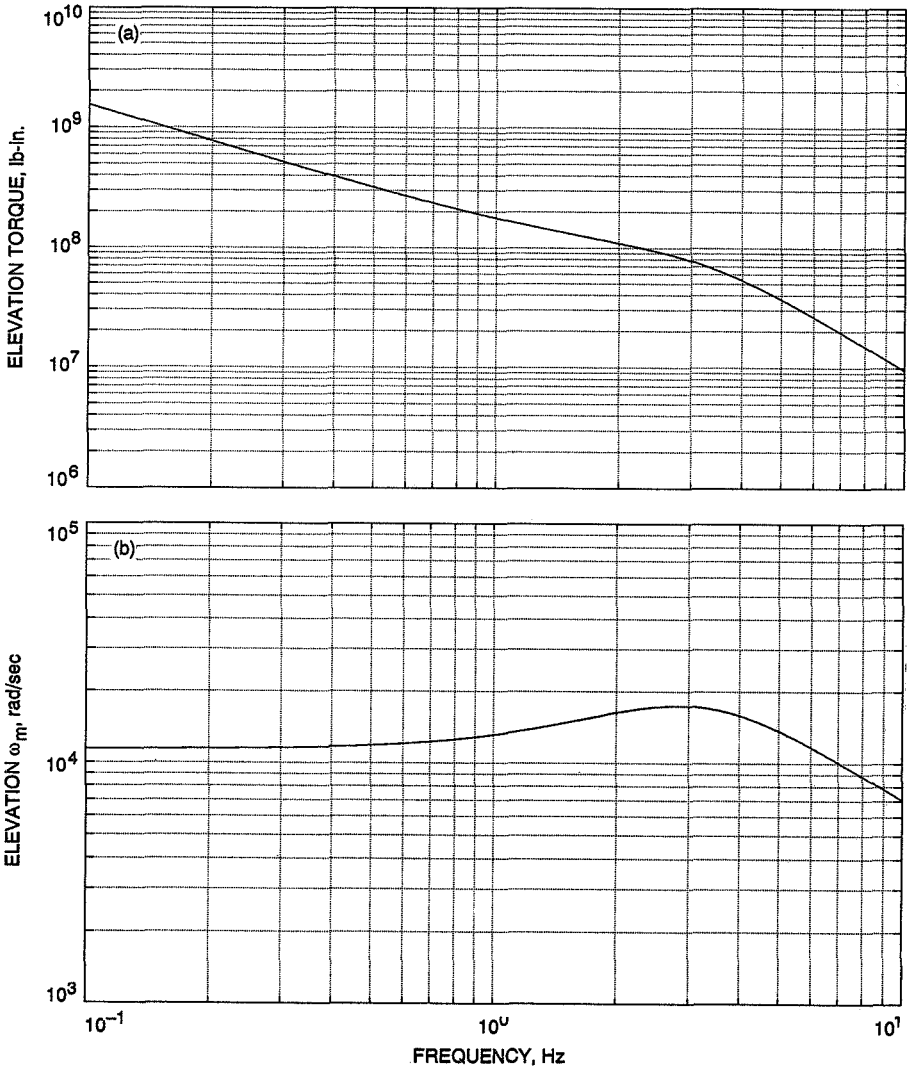


Fig. 7. Magnitudes of the transfer function of the full (16 states) and reduced (5 states) drive systems for the elevation drive input: rate command.

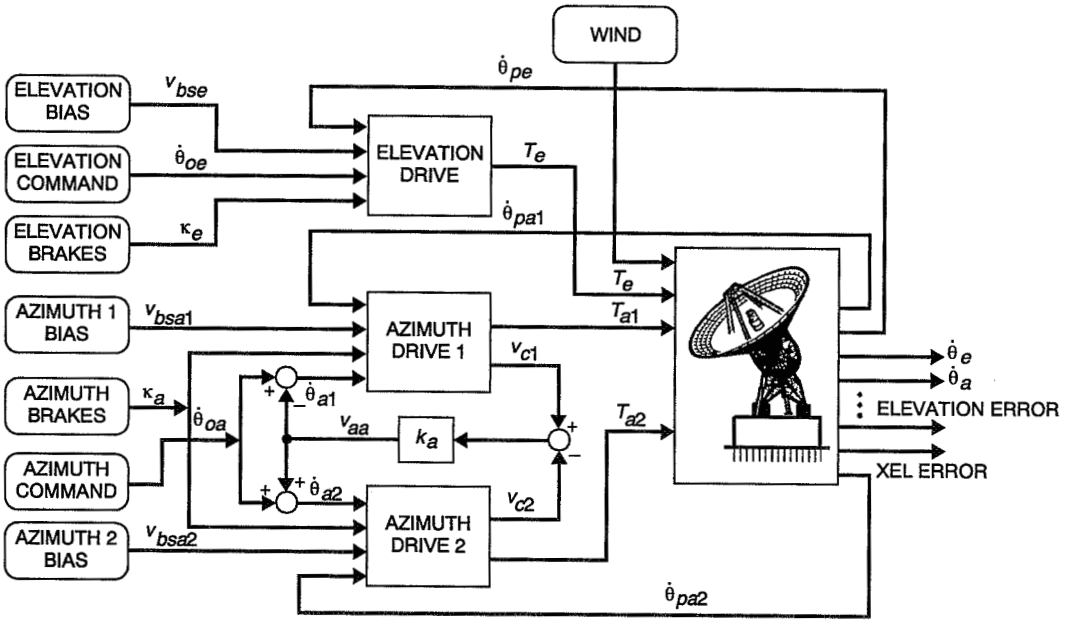


Fig. 8. Rate-loop model.

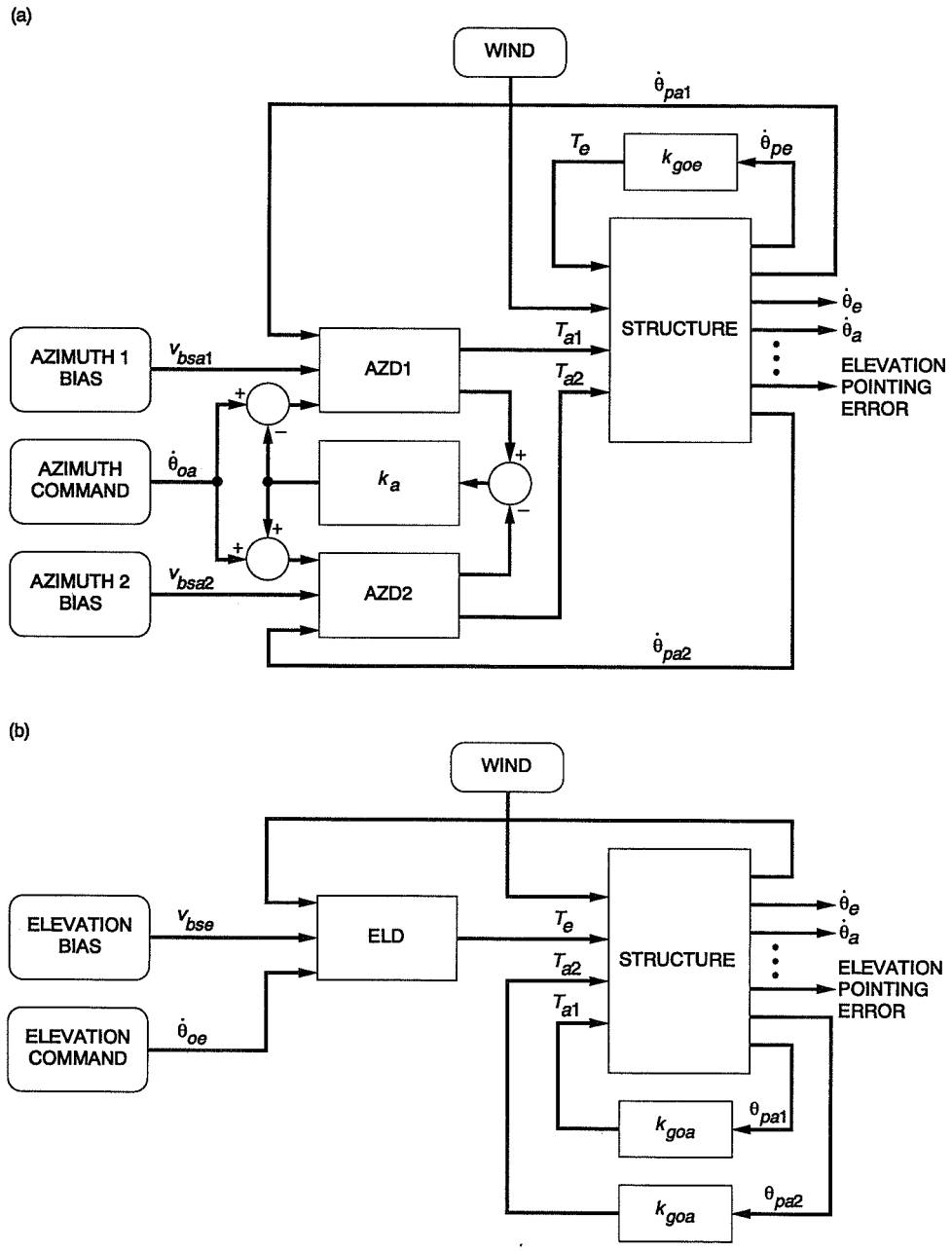


Fig. 9. Rate-loop model with active (a) elevation brakes and (b) azimuth brakes.

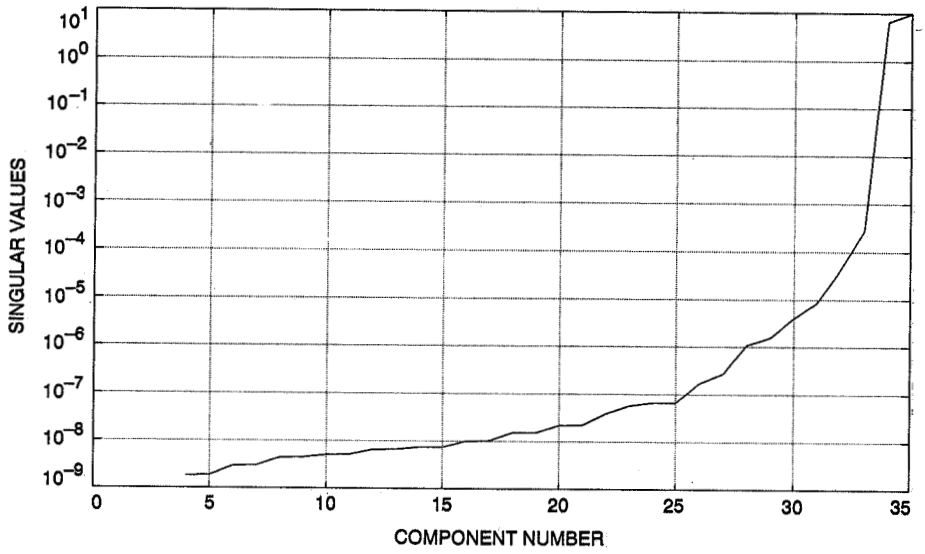


Fig. 10. Singular values of the balanced antigrammians for the rate-loop model.

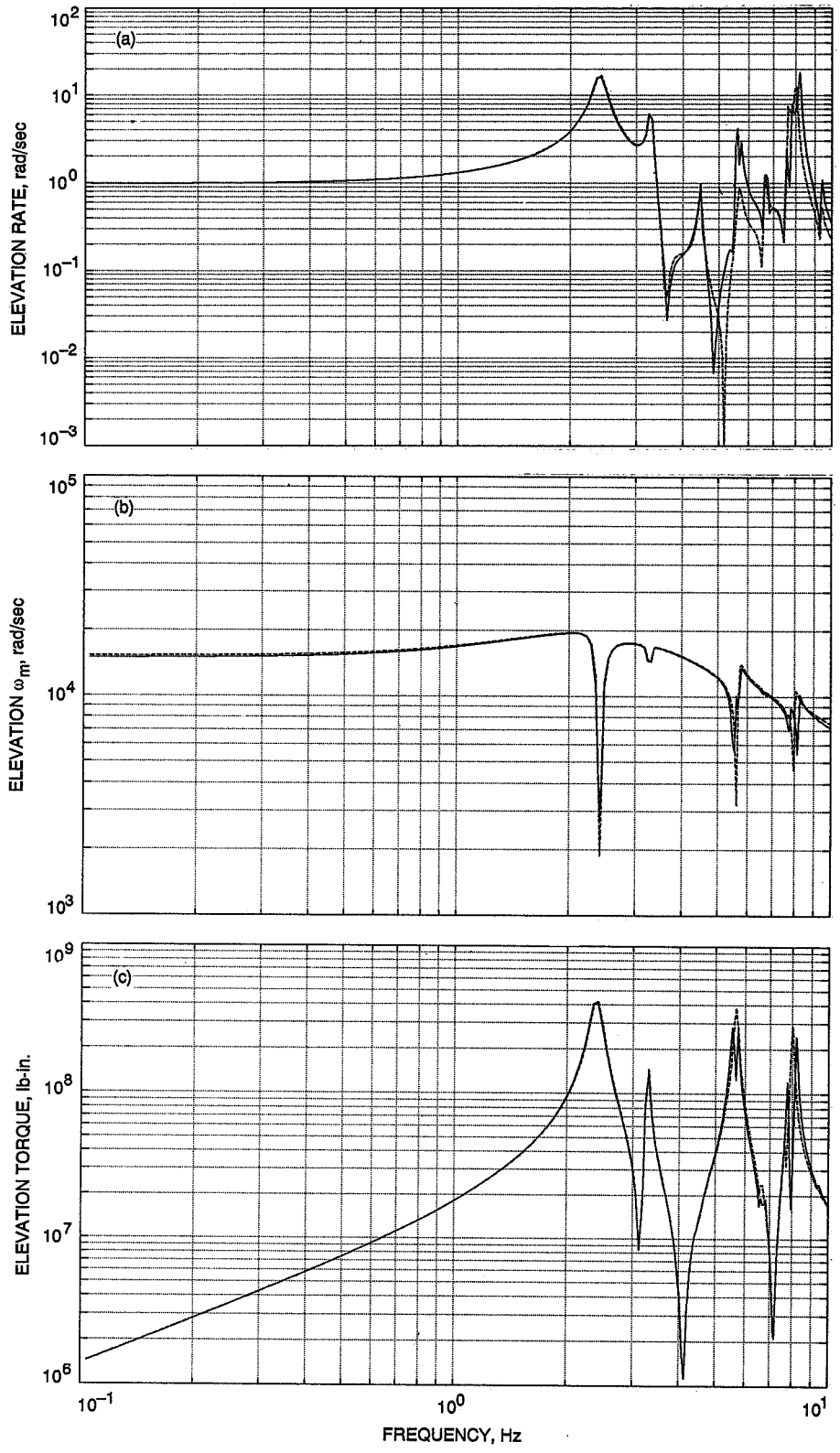


Fig. 11. Magnitudes of transfer function for the full (90 states) and the reduced (27 states) rate-loop model with elevation rate command input: (a) elevation rate, (b) elevation motor rate ω_m, (c) elevation torque, (d) azimuth rate, (e) AZ1 motor rate ω_m , and (f) AZ1 torque.

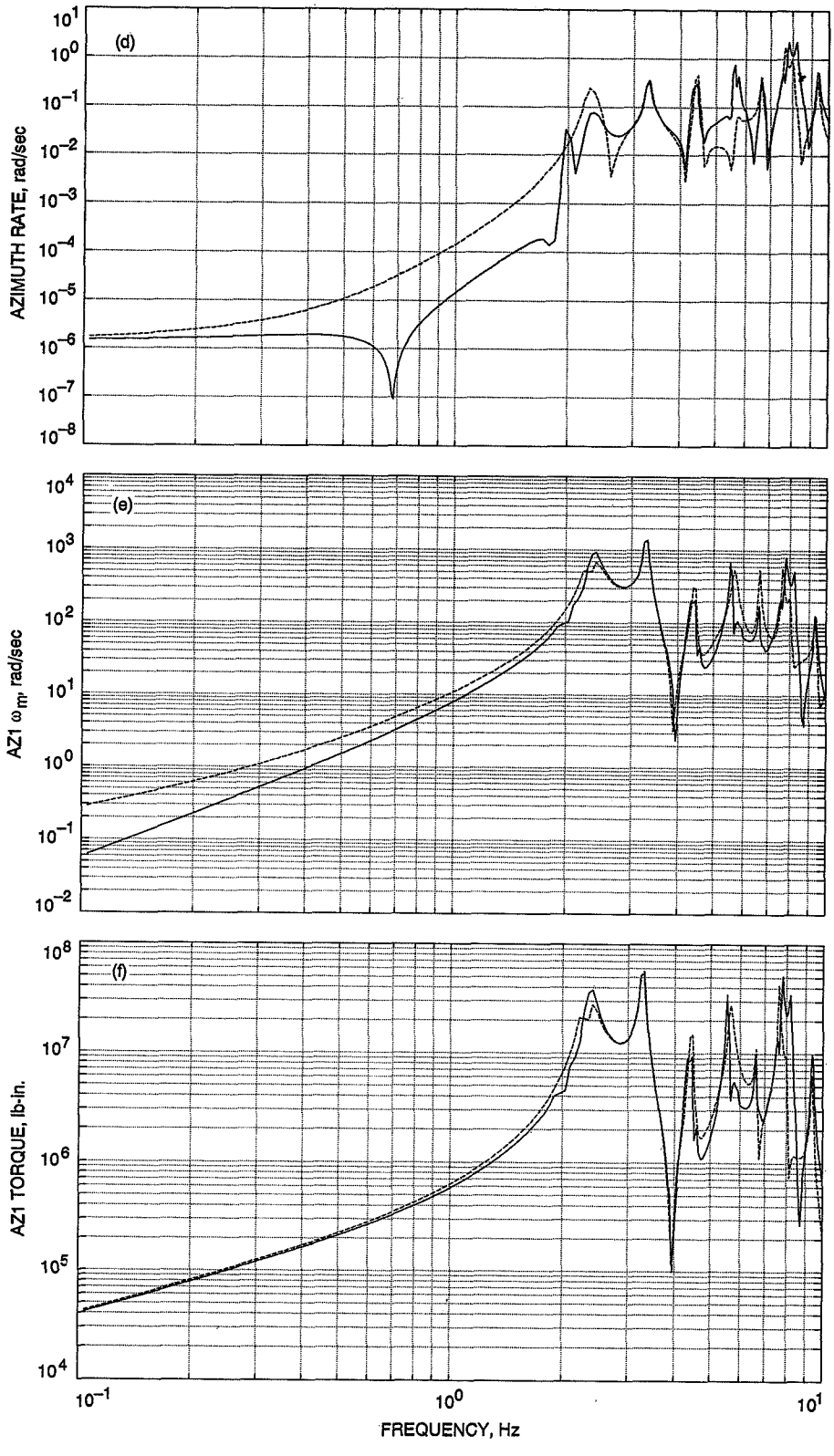

(d)
AZIMUTH RATE, rad/sec
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
(e)
AZ1 ω_m, rad/sec
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-2}
(f)
AZ1 TORQUE, lb-in.
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
10^{-1}
10^{0}
10^{1}
FREQUENCY, Hz

Fig. 11 (contd).

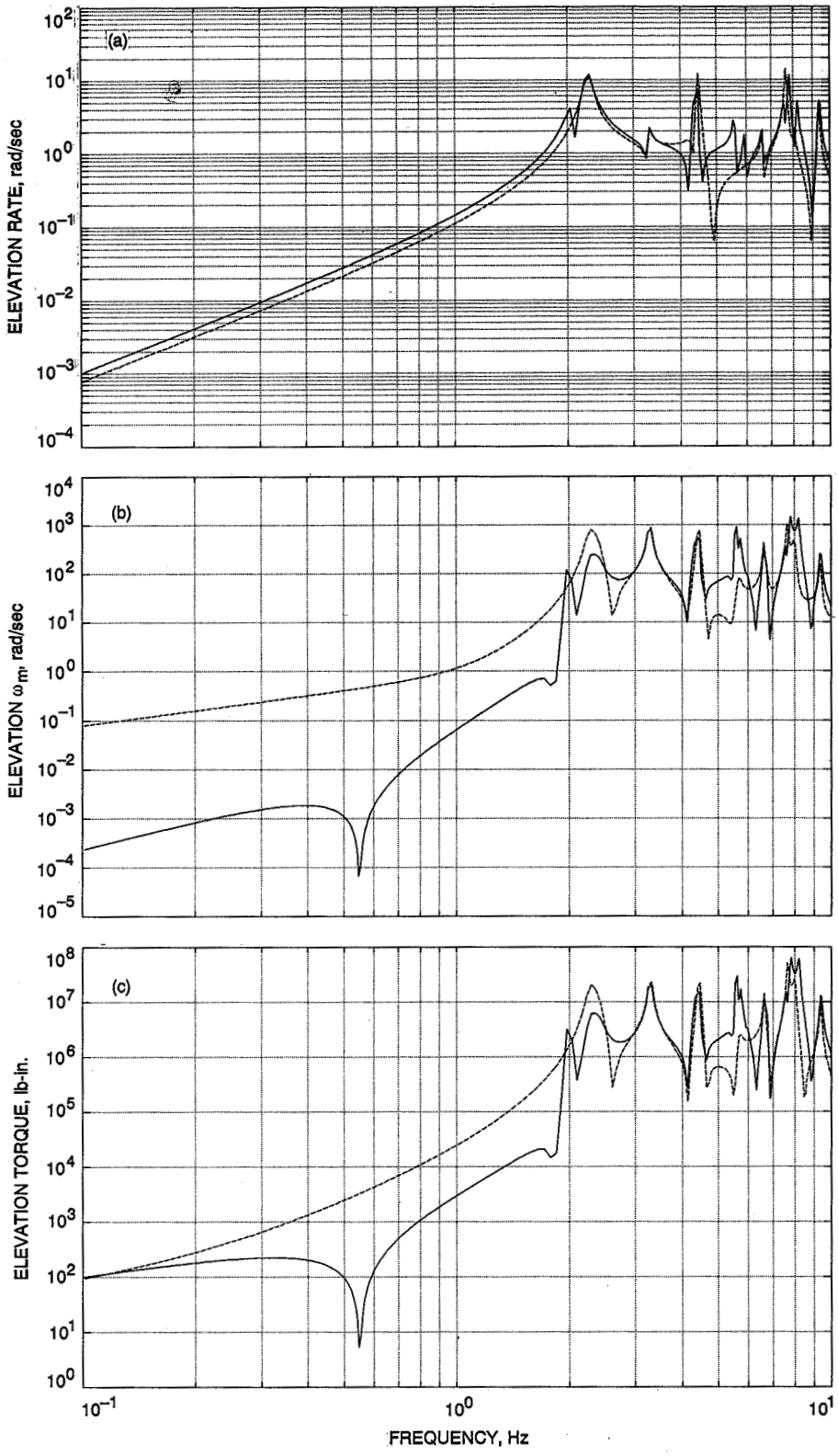


Fig. 12. Magnitudes of transfer function for the full (90 states) and the reduced (27 states) rate-loop model with azimuth rate command input: (a) elevation rate, (b) elevation motor rate ω_m , (c) elevation torque, (d) azimuth rate, (e) AZ1 motor rate ω_m , and (f) AZ1 torque.

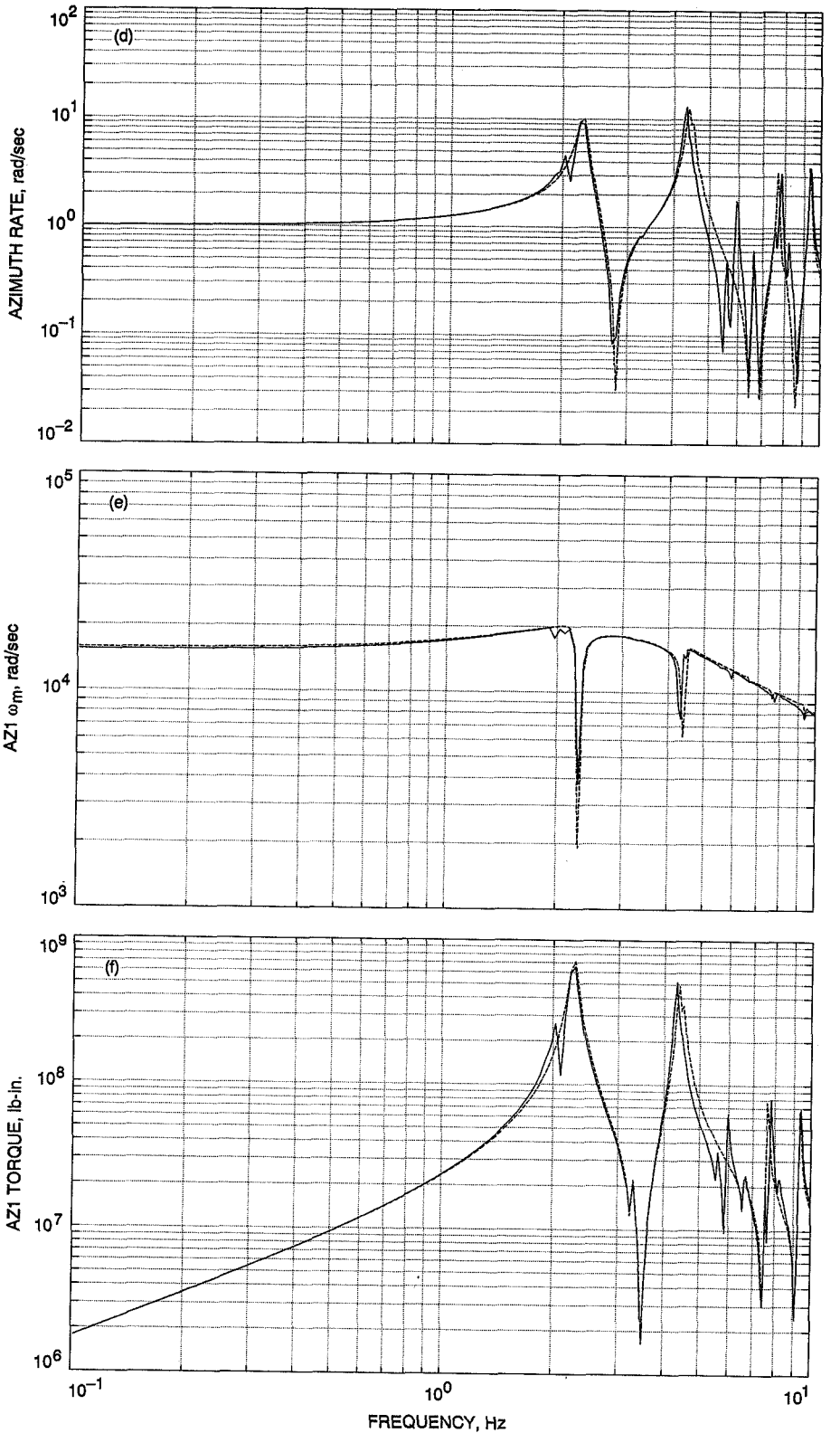

(d)
AZIMUTH RATE, rad/sec
10^2
10^1
10^0
10^{-1}
10^{-2}
(e)
AZ1 ω_m, rad/sec
10^5
10^4
10^3
(f)
AZ1 TORQUE, lb-in.
10^9
10^8
10^7
10^6
10^{-1}
10^0
10^1
FREQUENCY, Hz

Fig. 12 (contd).

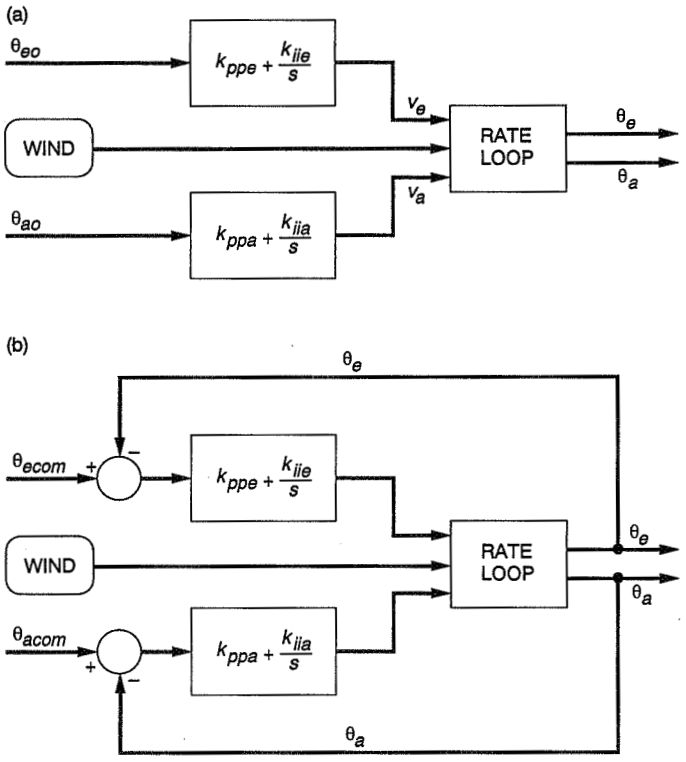


Fig. 13. Position loop system: (a) open and (b) closed.

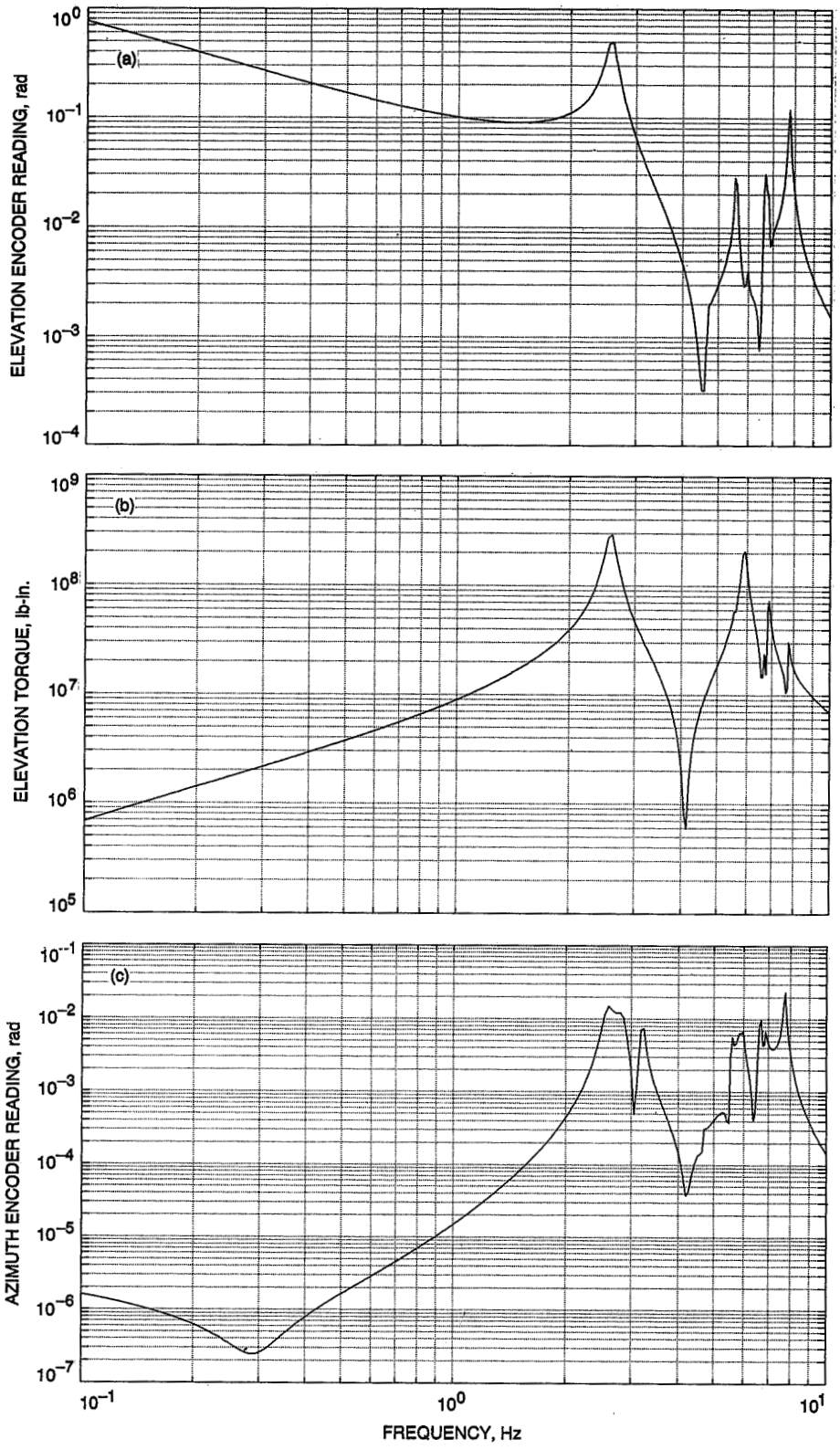


Fig. 14. Magnitudes of transfer function for the position loop system with elevation command: (a) elevation encoder reading, (b) elevation torque, (c) azimuth encoder reading, (d) AZ1 torque, (e) elevation pointing error, and (f) XEL pointing error.

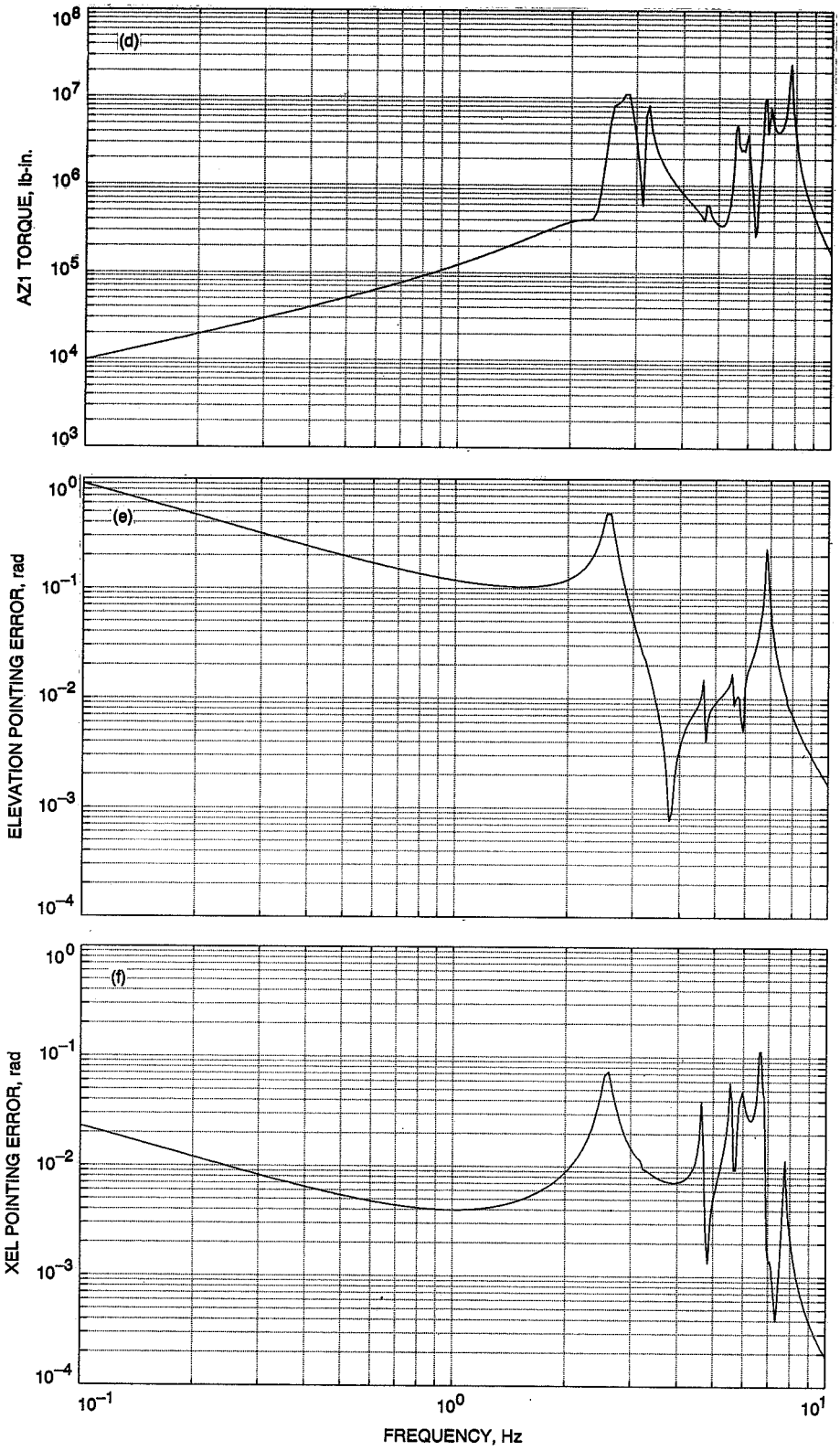

(d)
AZ1 TORQUE, lb-in.
10^8
10^7
10^6
10^5
10^4
10^3
(e)
ELEVATION POINTING ERROR, rad
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
(f)
XEL POINTING ERROR, rad
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-1}
10^0
10^1
FREQUENCY, Hz

Fig. 14 (contd).

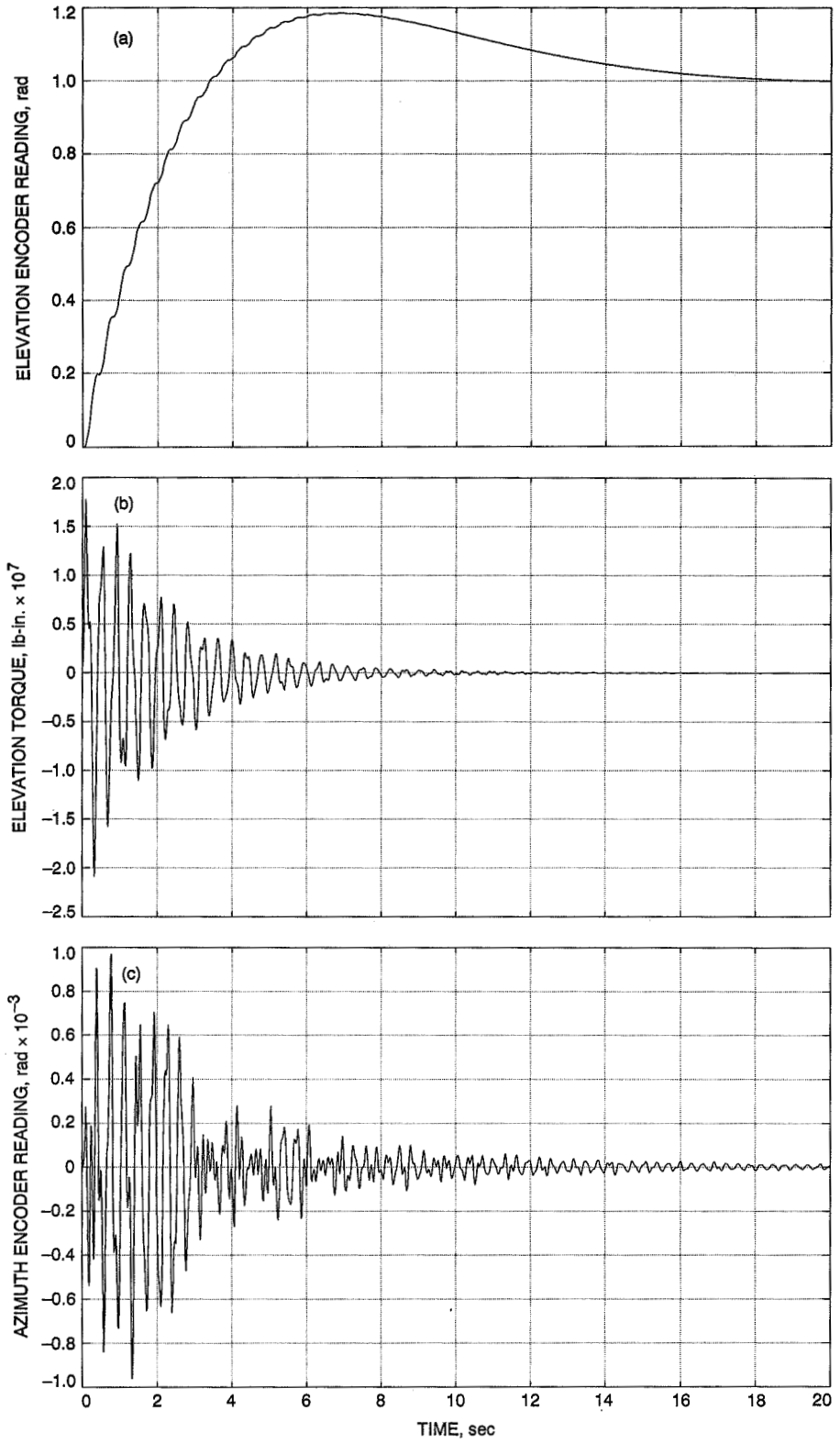


Fig. 15. Step responses of the position loop system with elevation command at 1 radian: (a) elevation encoder reading, (b) elevation torque, (c) azimuth encoder reading, (d) AZ1 torque, (e) elevation pointing error, and (f) XEL pointing error.

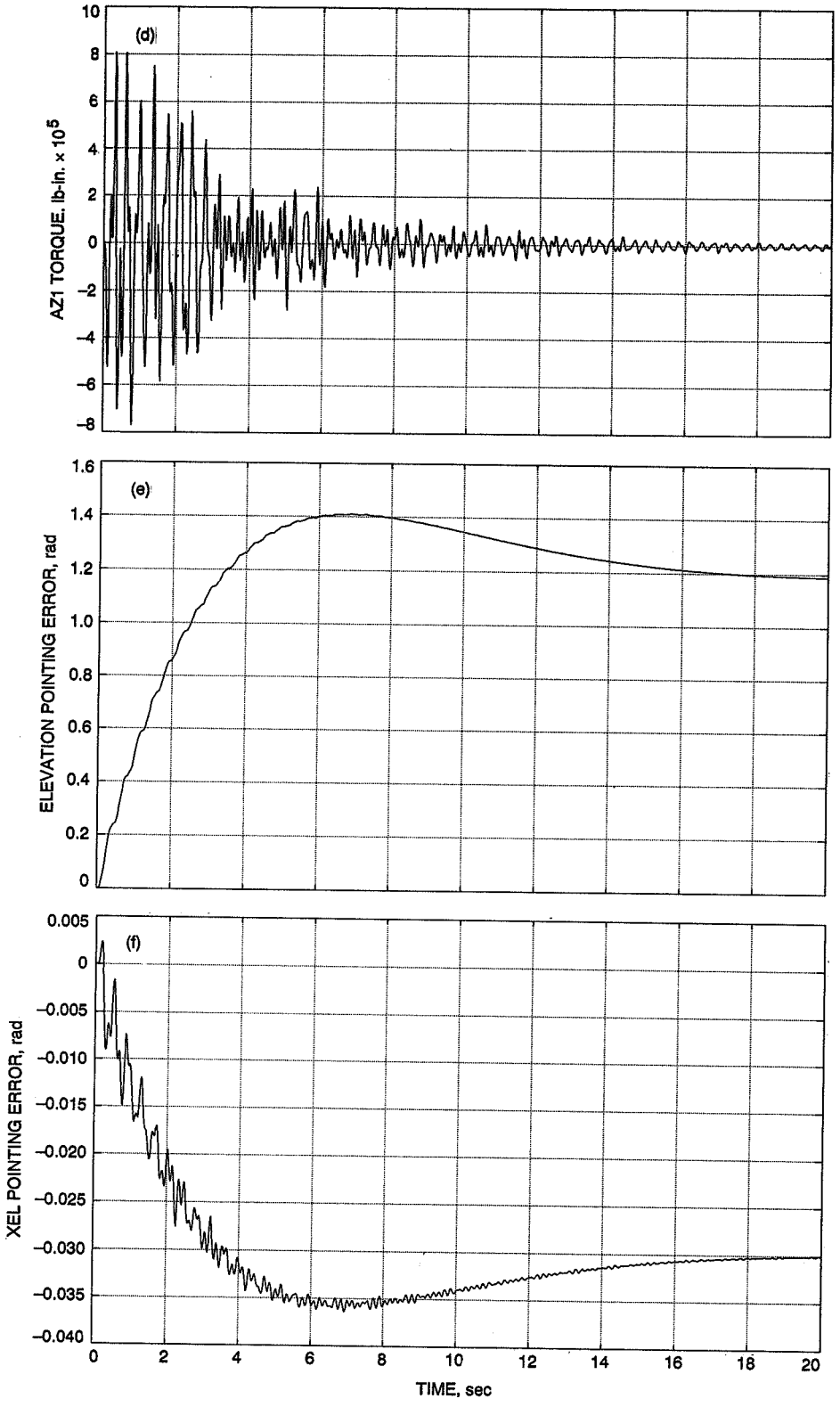

(d)
AZ1 TORQUE, lb-in. x 10^5
10
8
6
4
2
0
-2
-4
-6
-8
(e)
ELEVATION POINTING ERROR, rad
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
(f)
XEL POINTING ERROR, rad
0.005
0
-0.005
-0.010
-0.015
-0.020
-0.025
-0.030
-0.035
-0.040
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 15 (contd).

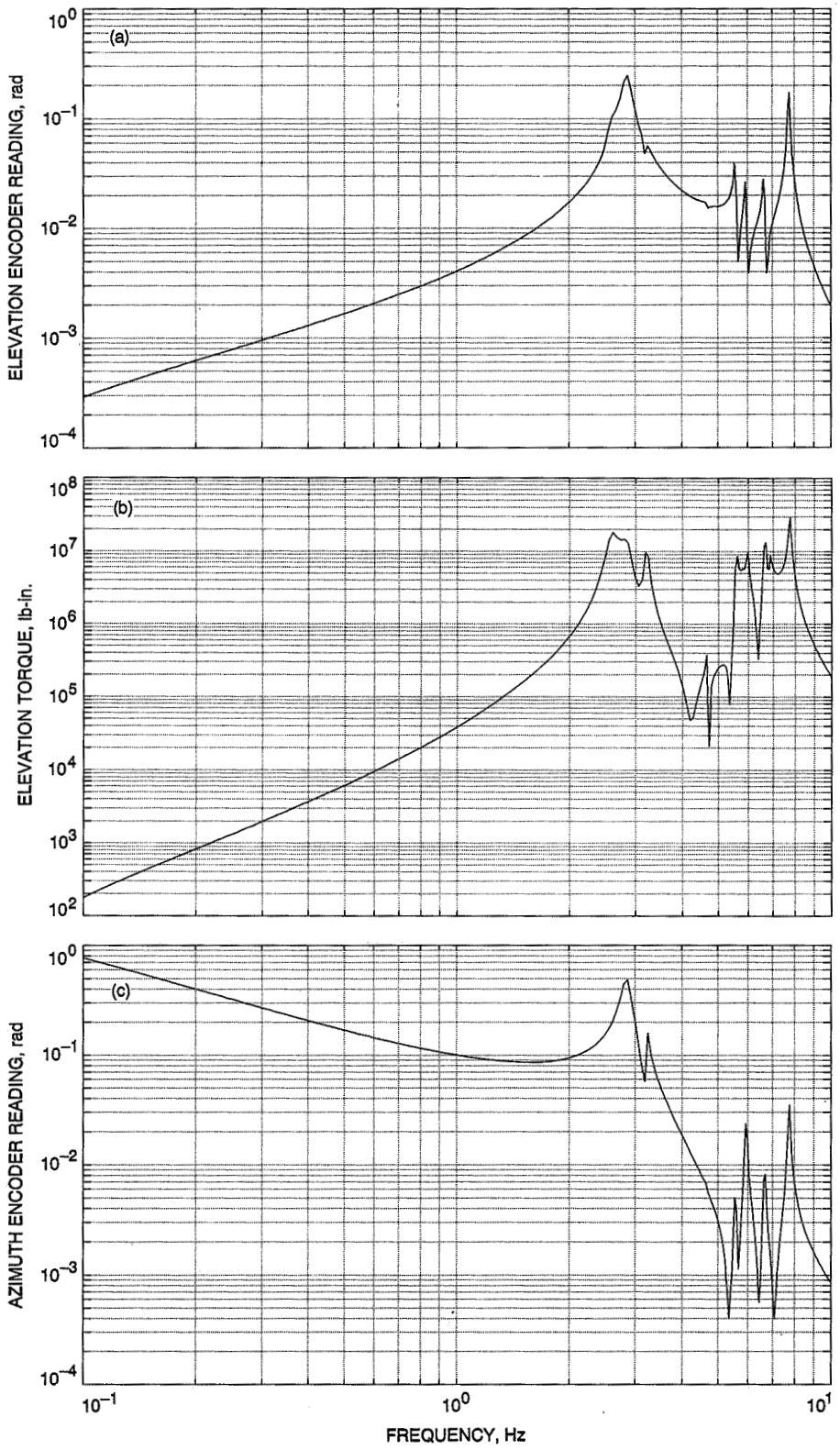


Fig. 16. Magnitudes of transfer function for the position loop system with azimuth command: (a) elevation encoder reading, (b) elevation torque, (c) azimuth encoder reading, (d) AZ1 torque, (e) elevation pointing error, and (f) XEL pointing error.

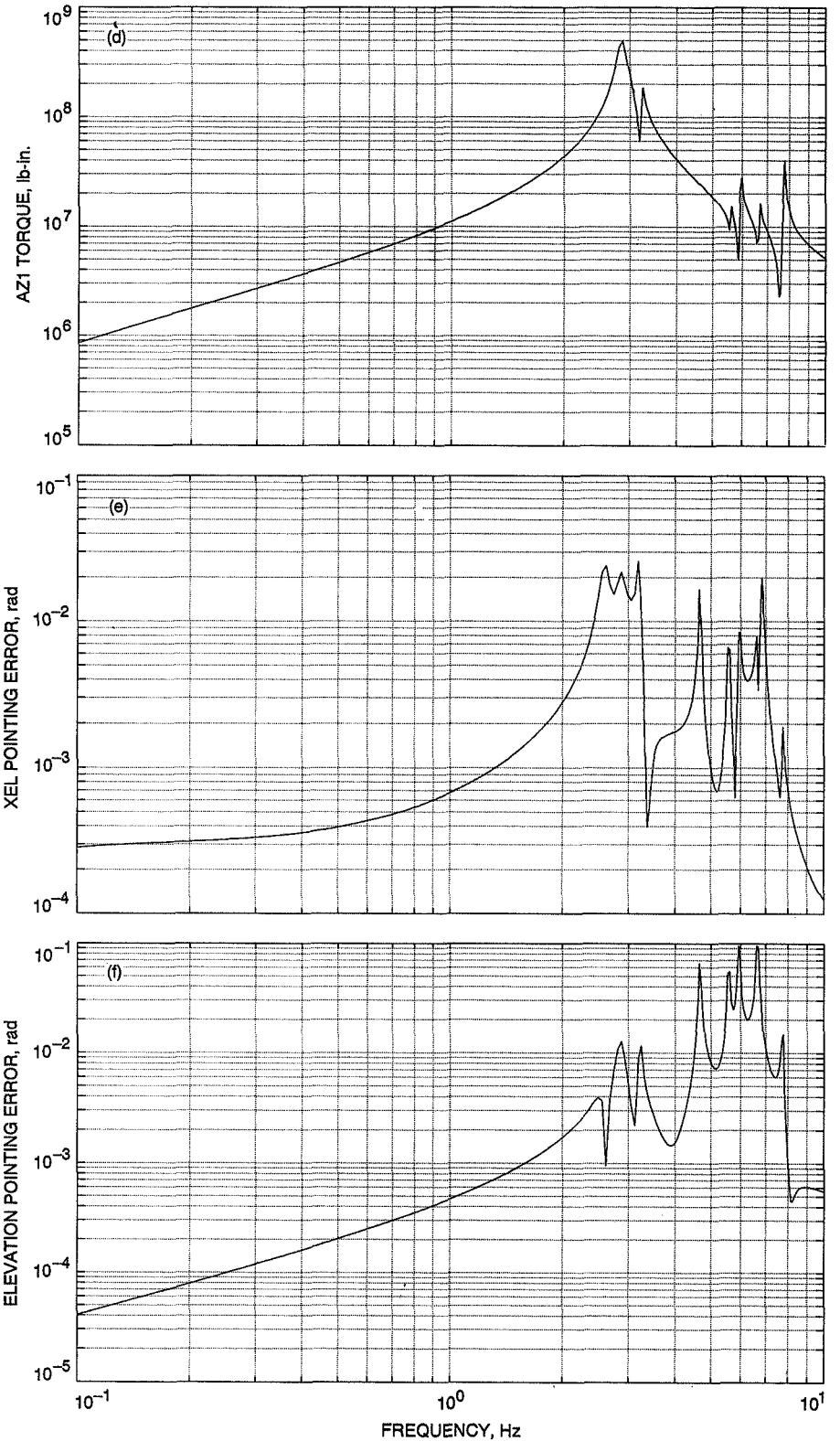

(d)
AZ1 TORQUE, lb-in.
10^9
10^8
10^7
10^6
10^5
(e)
XEL POINTING ERROR, rad
10^{-1}
10^{-2}
10^{-3}
10^{-4}
(f)
ELEVATION POINTING ERROR, rad
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-1}
10^0
10^1
FREQUENCY, Hz

Fig. 16 (contd).

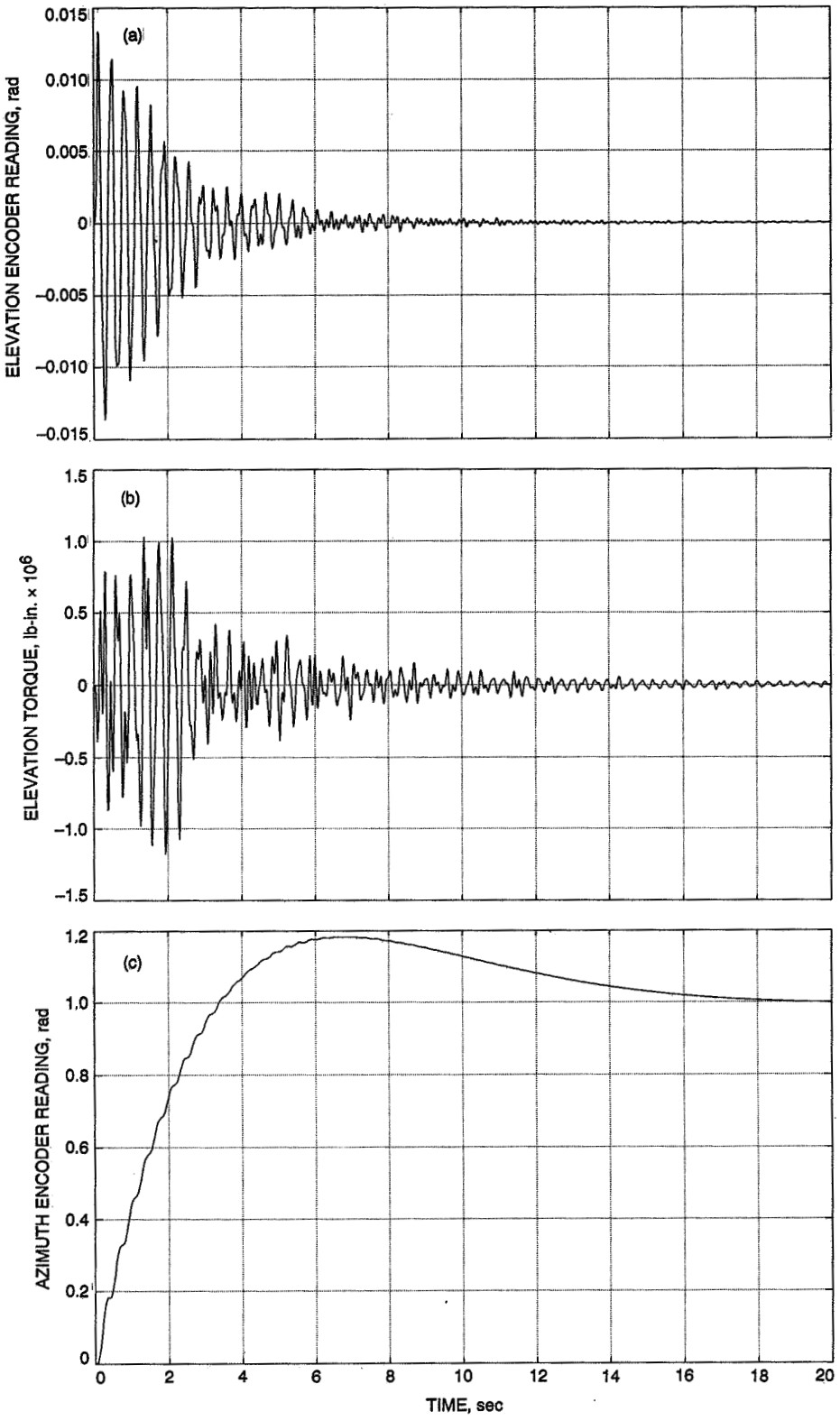


Fig. 17. Step responses of the position loop system with azimuth command at 1 radian: (a) elevation encoder reading, (b) elevation torque, (c) azimuth encoder reading, (d) AZ1 torque, (e) elevation pointing error, and (f) XEL pointing error.

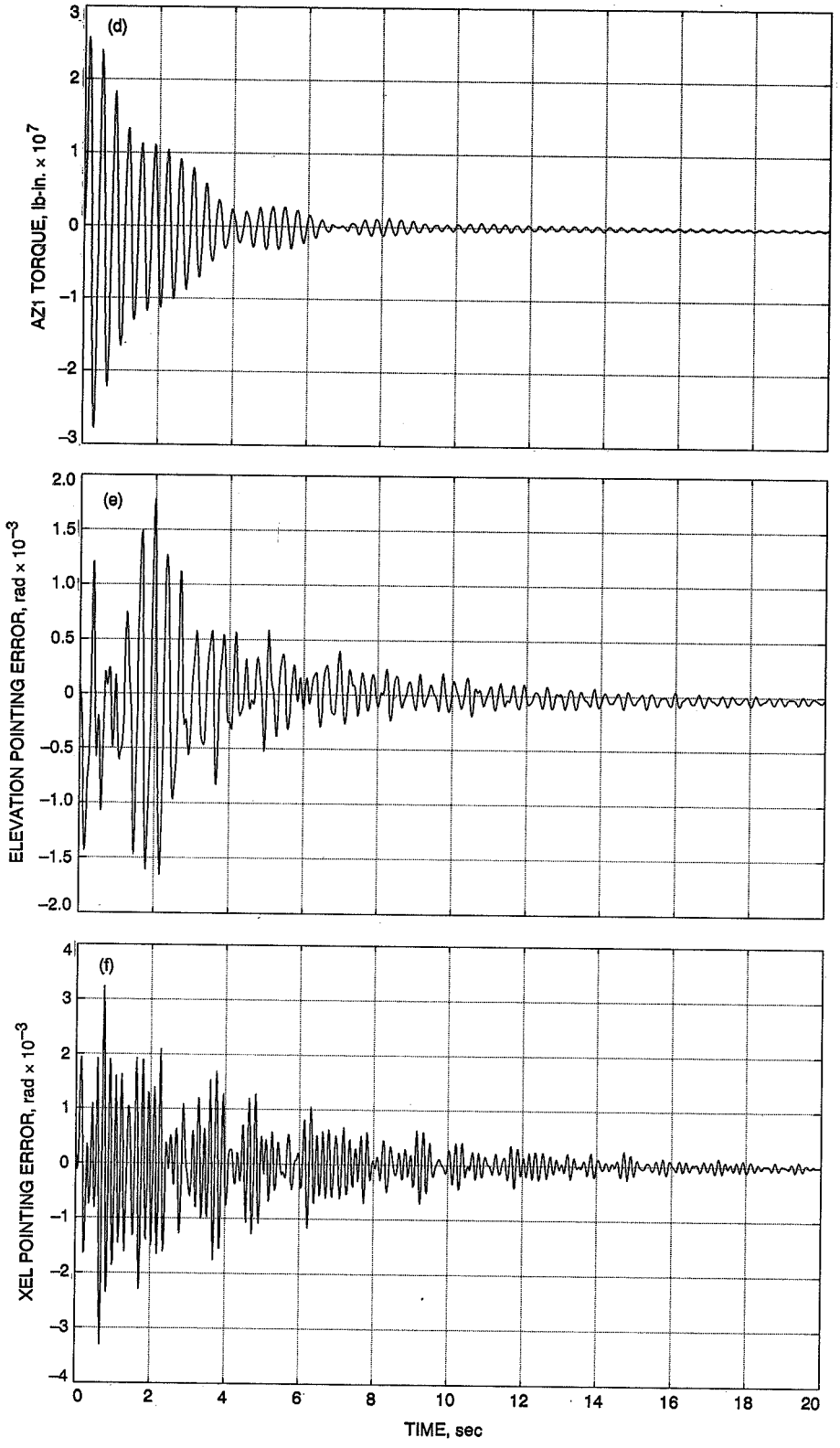

(d)
AZ1 TORQUE, lb-in. x 10^7
3
2
1
0
−1
−2
−3
(e)
ELEVATION POINTING ERROR, rad x 10^−3
2.0
1.5
1.0
0.5
0
−0.5
−1.0
−1.5
−2.0
(f)
XEL POINTING ERROR, rad x 10^−3
4
3
2
1
0
−1
−2
−3
−4
0
2
4
6
8
10
12
14
16
18
20
TIME, sec

Fig. 17 (contd).

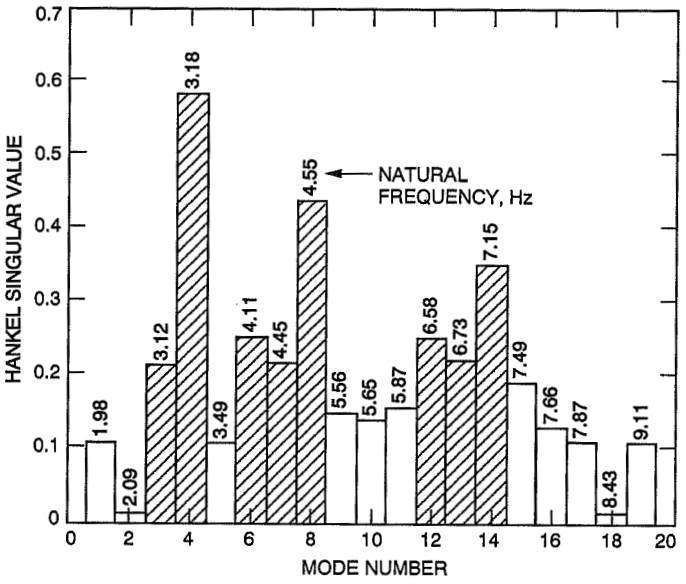


Fig. 18. Hankel singular values for the antenna structure with the wind input.

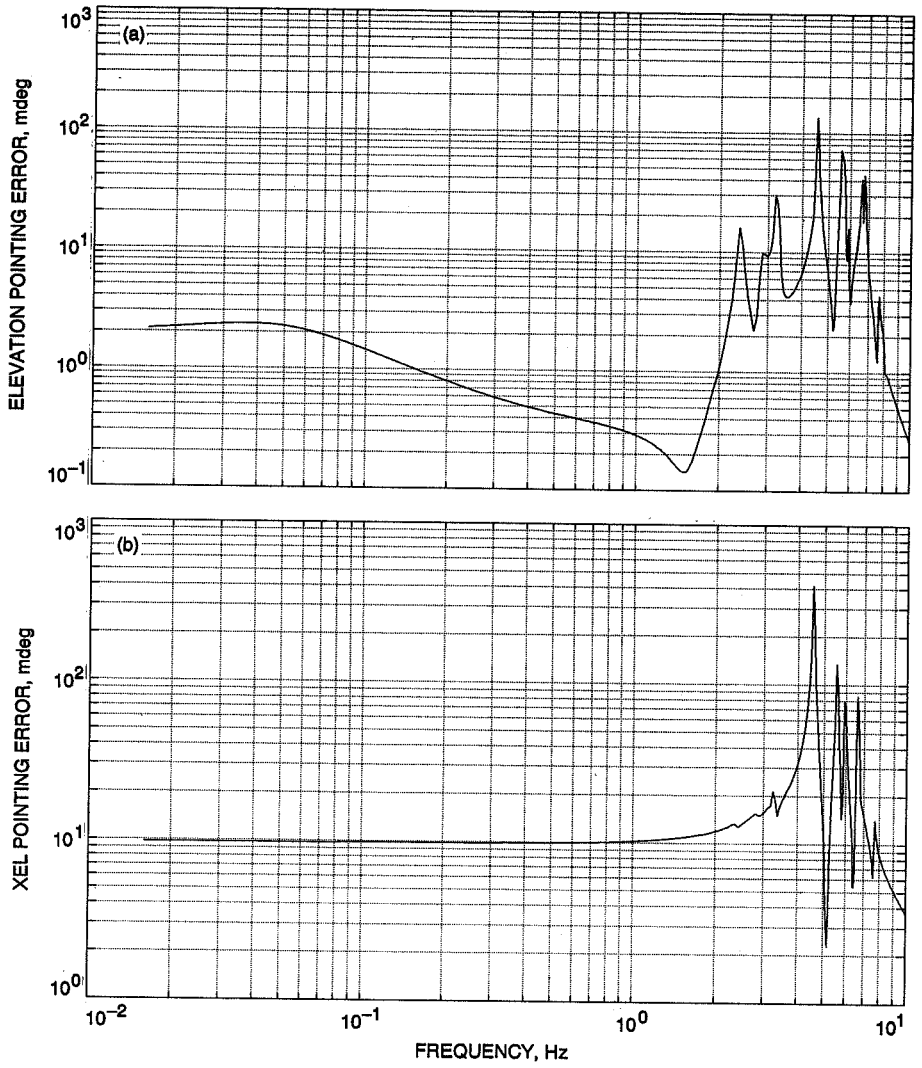


Fig. 19. Magnitudes of transfer functions of the closed-loop system: (a) from the wind force in *x*-direction to the elevation pointing error, (b) from the wind force in *x*-direction to the XEL pointing error, (c) from the wind force in *y*-direction to the elevation pointing error, and (d) from the wind force in *y*-direction to XEL pointing error.

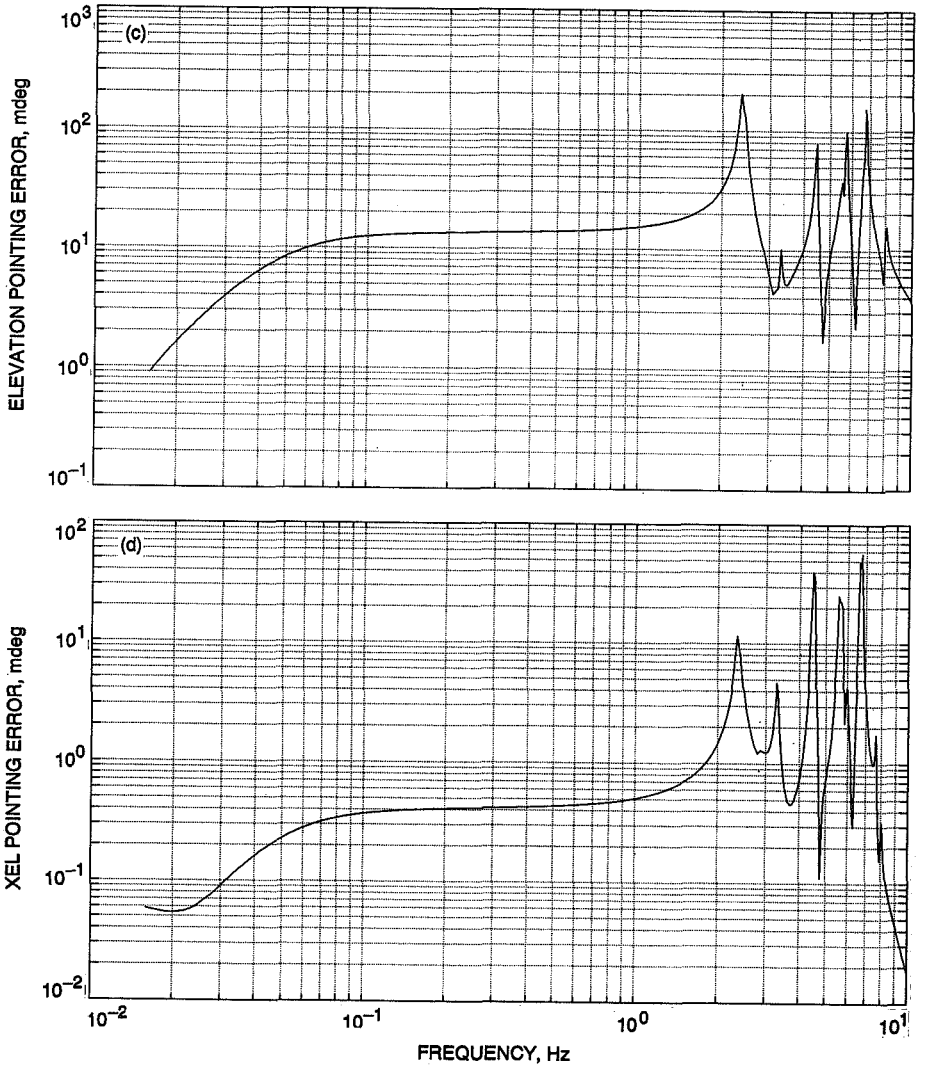
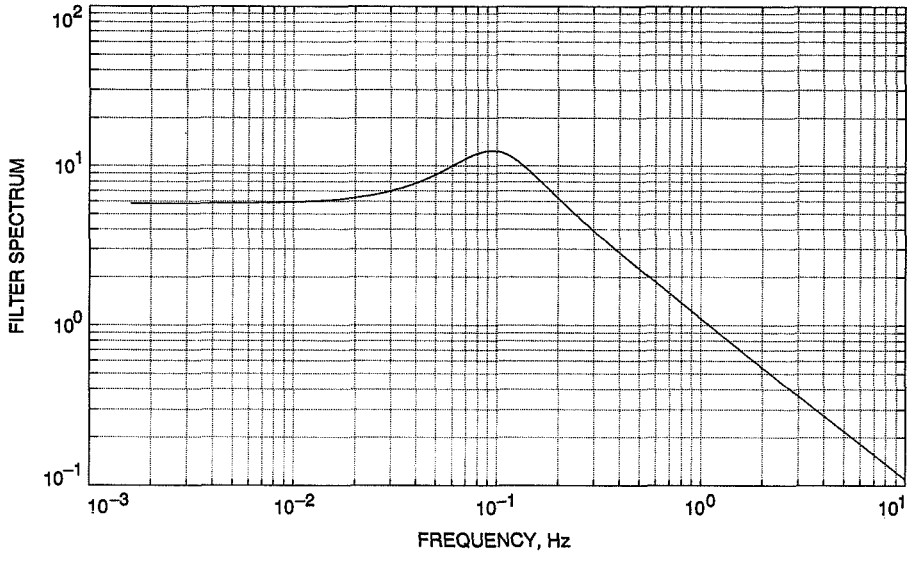

(c)
ELEVATION POINTING ERROR, mdeg
10^3
10^2
10^1
10^0
10^{-1}
(d)
XEL POINTING ERROR, mdeg
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-2}
10^{-1}
10^0
10^1
FREQUENCY, Hz

Fig. 19 (contd).



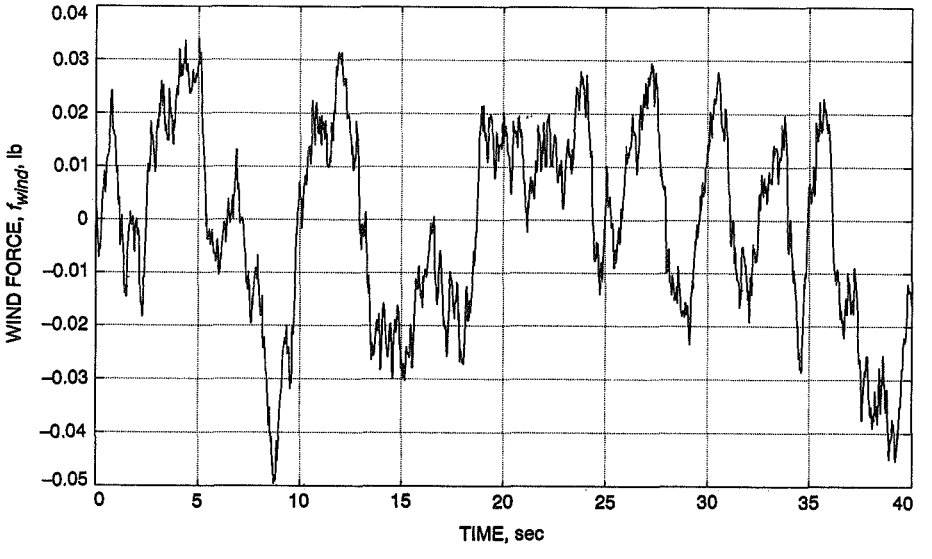


Fig. 21. Wind force acting on the tipping structure for wind 30 mph and with $f_{max} = 0.1$ Hz and rms (f_{wind}) = 0.18.

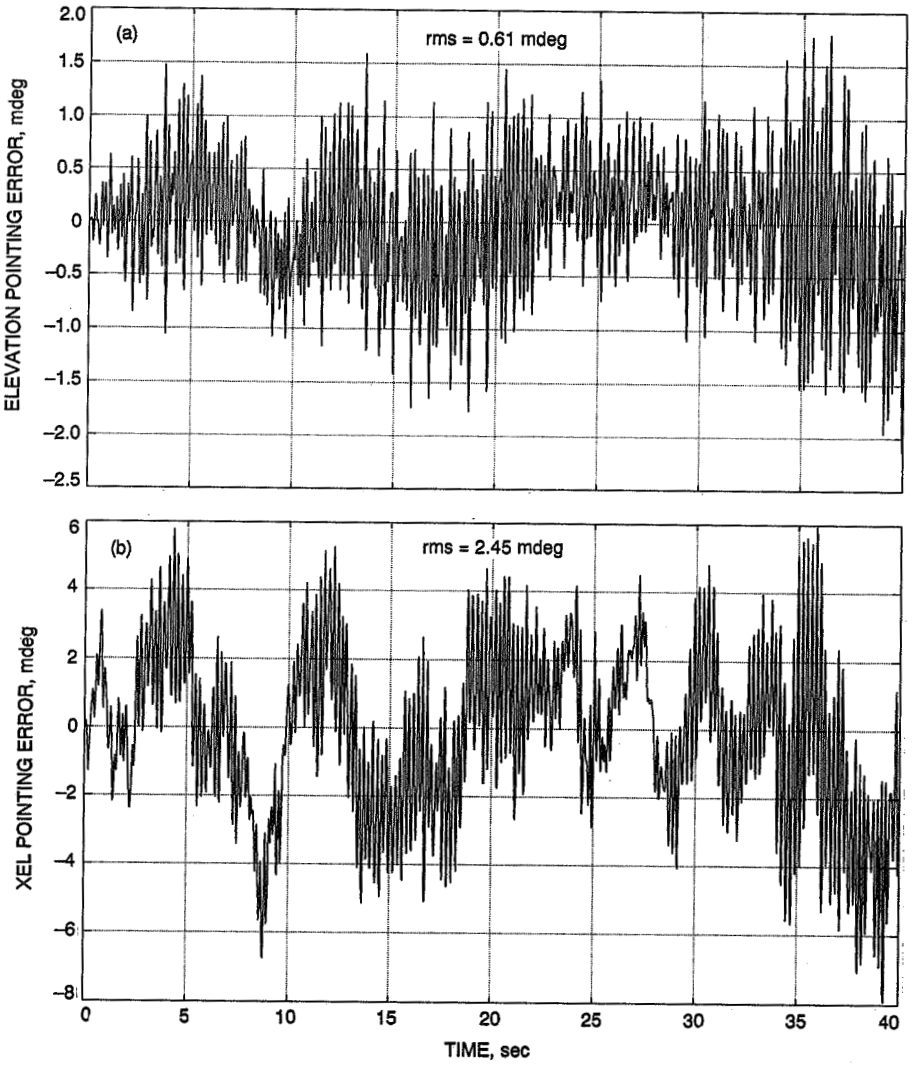


Fig. 22. Closed-loop antenna responses to 30 mph wind, $f_{max} = 0.1$ Hz, and rms (f_{wind}) = 0.18: (a) elevation pointing error, *x*-direction wind, (b) XEL pointing error, *x*-direction wind, (c) elevation pointing error, *y*-direction wind, and (d) XEL pointing error, *y*-direction wind.

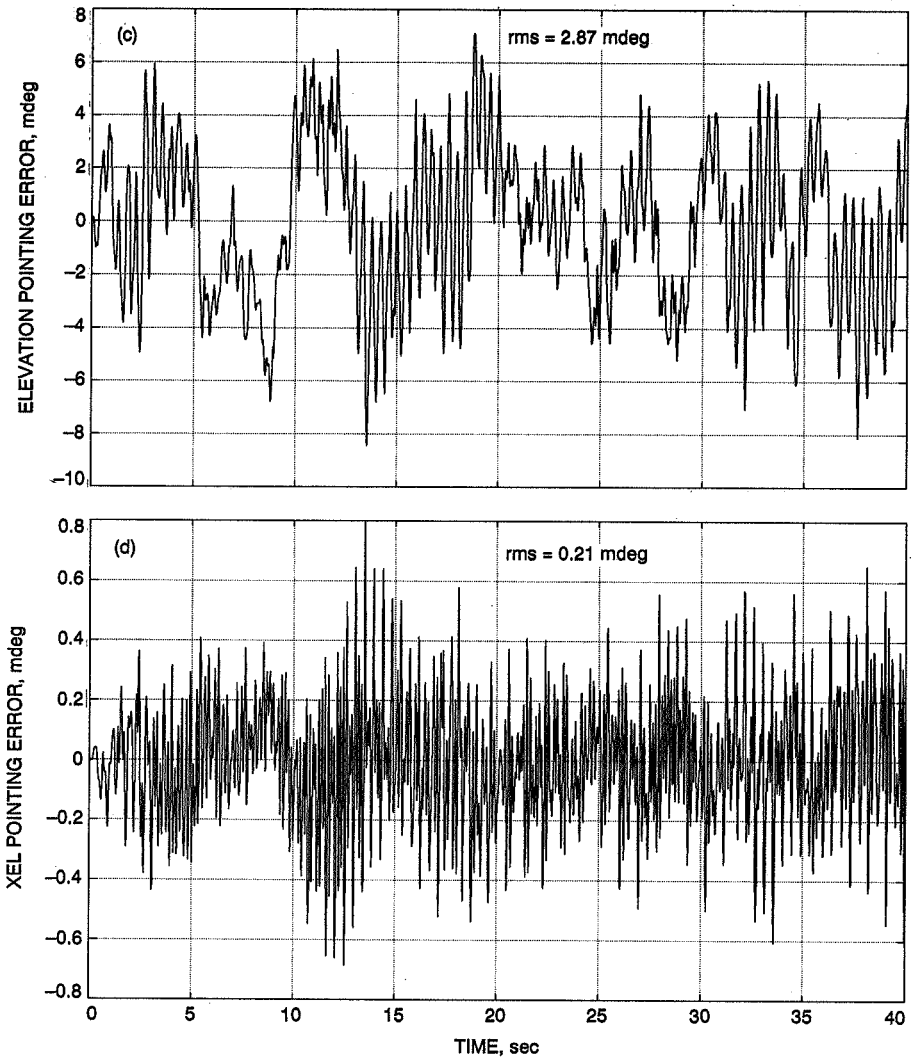

(c)
rms = 2.87 mdeg
ELEVATION POINTING ERROR, mdeg
8
6
4
2
0
−2
−4
−6
−8
−10
(d)
rms = 0.21 mdeg
XEL POINTING ERROR, mdeg
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
0
5
10
15
20
25
30
35
40
TIME, sec

Fig. 22 (contd).

Appendix A
Parameters of the DSS 13 Antenna

The structural parameters for the DSS 13 antenna are:

$\alpha = 25$ deg

$\zeta_i = 0.005$ for $i = 1, \cdots, 21$

node numbers are $nc = 5380, no = 86,302, nu = 86,881$, and $nb = 41,212$

$$\begin{aligned}\Omega = diag\ (\ & 0, & & 0, & & 12.4344, & & 13.1067, & & 19.6287, & & 19.9805, & & 21.9158,\\ & 25.8490, & & 27.9288, & & 28.6011, & & 34.9094, & & 35.5126, & & 36.8760, & & 41.3559,\\ & 42.3047, & & 44.9499, & & 47.0485, & & 48.1292, & & 49.4675, & & 52.9547, & & 57.2524)\end{aligned}$$

$$\begin{aligned}M_m = diag\ (\ & 214.3607, & & 296.0889, & & 692.2613, & & 11.3445, & & 244.8719, & & 62.6765, & & 65.7475,\\ & 799.5434, & & 17.9187, & & 18.6408, & & 18.9721, & & 24.6747, & & 53.4131, & & 7.8255,\\ & 9.7606, & & 37.4093, & & 38.2110, & & 27.3915, & & 25.5052, & & 7.5668, & & 200.0856)\end{aligned}$$

modal matrix $\Phi = [\phi_1, \phi_2, \cdots, \phi_{21}]$, where $\phi_i = i$th mode shape and $\phi_i = [\phi_{i1}, \phi_{i2}, \cdots, \phi_{ip}]^T$, where $\phi_{ij} = j$th component of the ith mode.

The rate-loop system parameters for the DSS 13 antenna are:

$k_1 = 716.197$ V sec/rad

$k_m = 15.72$ lb/A for elevation

$k_m = 15.36$ lb/A for azimuth

$k_b = 1.79$ V sec/rad

$k_s = 0.8$ V/V

$k_{tach} = 0.0384123$ V sec/rad for elevation

$k_{tach} = 0.037337$ V sec/rad for azimuth

$k_r = 80$ V/sec V for a range of 49–83

$k_i = 87.13$ V/sec V

$k_{iie} = k_{iia} = 0.1$

$k_{ppe} = k_{iia} = 0.5$

$k_{cur} = 0.12658$ V/A

$k_f = 54$ V/V

$k_{go} = 1.5 \times 10^7$ lb/rad for elevation

$k_{go} = 2.0 \times 10^7$ lb/rad for azimuth

$k_{ctfr} = 0.33$

$k_{ct} = 0.11111$

$k_{bs} = 0.66$

$\tau_1 = 0.0063662$ sec

$\tau_2 = 0.094$ sec

$\tau_3 = 0.002$ sec

$\tau_4 = 0.00484$ sec

$\tau_5 = 0.0021$ sec

$\tau_6 = 0.7304$ sec

$N = 354$ for elevation

$N = 595$ for azimuth

$J_m = 1.236$ lb/sec^2 for elevation

$J_m = 1.0848$ lb/sec^2 for azimuth

$R_a = 0.456\ \Omega$ for elevation

$R_a = 0.668\ \Omega$ for azimuth

$L_a = 0.011\ H$ for elevation

$L_a = 0.0144\ H$ for azimuth

Appendix B

Model Reduction for Systems With Integrators

Model order reduction methods for stable linear systems are based on joint controllability and observability tests, through balancing of system grammians [2]. For linear unstable systems, various techniques based on properties of the controllability and observability grammians have been developed [8,9,10]. Consider a system with integrators whose output is an integral of an input. It has a pole (or poles) at zero, consequently, its controllability and observability grammians do not exist, and model reduction based on its grammian properties cannot be executed. Systems with integrators, however, are controllable and observable, hence, these properties can still be used for model reduction. In this appendix the reduction algorithm for systems with integrators is derived, and systems with integrators are defined as follows.

Definition B.1. A linear system with the state-space representation (A, B, C, D), $A \in \Re^{n\times n}$, $B \in \Re^{n\times p}$, $C \in \Re^{q\times n}$, and $D \in \Re^{q\times p}$, is a system with integrators if its $n - m$ poles are stable, the remaining m poles are at zero, and it is observable and controllable, see [11]. It is assumed also that A is nondefective [12] (geometric multiplicity of poles at zero is m).

Systems with integrators are linear systems commonly encountered in control design and analysis. This appendix presents a reduction method, by introducing antigrammians.

Definition B.2. For a controllable and observable triple (A, B, C), the matrices V_c and V_o satisfying the following Riccati equations:

$$\left.\begin{aligned} V_cA + A^TV_c + V_cBB^TV_c = 0 \\ V_0A^T + AV_o + V_oC^TCV_o = 0 \end{aligned}\right\} \qquad \text{(B-1a)}$$

are the controllability and observability antigrammians.

For stable controllable and observable systems $V_o = W_c^{-1}$ and $V_o = W_o^{-1}$, where W_c and W_o are controllability and observability grammians and satisfy the Lyapunov equations

$$\left.\begin{aligned} AW_c + W_cA^T + BB^T = 0 \\ W_oA + A^TW_o + C^TC = 0 \end{aligned}\right\} \qquad \text{(B-1b)}$$

Note that the grammians for a system with integrators do not exist, but antigrammians do; for an unobservable or uncontrollable system antigrammians do not exist, but grammians do. The existence of antigrammians is exploited for the balancing and model reduction of systems with integrators.

For a stable, controllable, observable, and balanced system the grammians as well as the antigrammians are equal and diagonal

$$W_c = W_o = \Gamma, V_c = V_o = \Pi, \qquad \Pi = \Gamma^{-1} \qquad \text{(B-2)}$$

where $\Gamma = diag(\gamma_i)$ and $\Pi = diag(\pi_i)$ for $i = 1, 2, \ldots, n$, and satisfy the following equations:

$$\left.\begin{aligned} \Pi A_b + A_b^T\Pi + \Pi B_bB_b^T\Pi = 0 \\ \Pi A_b^T + A_b\Pi + \Pi C_b^TC_b\Pi = 0 \end{aligned}\right\} \qquad \text{(B-3a)}$$

$$\left.\begin{aligned} A_b\Gamma + \Gamma A_b^T + B_bB_b^T = 0 \\ \Gamma A_b + A_b^T\Gamma + C_b^TC_b = 0 \end{aligned}\right\} \qquad \text{(B-3b)}$$

The representation (A_b, B_b, C_b, D) is balanced.

Proposition B.1. For a balanced system with integrators one obtains $\Pi = diag(0_m\ \Pi_o)$, where 0_m is an $m \times m$ zero matrix, and A_b is block-diagonal, $A_b = diag(0_m\ A_{bo})$.

Proof. Consider A, B, and C in the form

$$A = diag(0_m\ A_o), \quad B^T = [B_r^T\ B_o^T], \quad C = [C_r\ C_o] \qquad \text{(B-4)}$$

Matrix A in the form of Eq. (B-4) always exists due to m poles at zero, and B and C exist due to nondefectiveness of A. From Eq. (B-1a) it follows that:

$$V_{crr}B_r = 0, \quad C_rV_{orr} = 0, \quad V_{cro} = 0, \quad V_{oro} = 0 \qquad \text{(B-5)}$$

where V_c and V_o are divided conformably to A:

$$V_c = \begin{bmatrix} V_{crr} & V_{cro} \\ V_{cro}^T & V_{coo} \end{bmatrix}, \qquad V_o = \begin{bmatrix} V_{orr} & V_{oro} \\ V_{oro}^T & V_{ooo} \end{bmatrix}$$

For a controllable and observable system (by Definition B.1) the matrices B_r and C_r are of full rank, thus it follows from Eq. (B-5) that $V_{crr} = 0, V_{orr} = 0$, and that $V_c = diag(0_m\ V_{coo}), V_o = diag(0_m\ V_{ooo})$, which in balanced coordinates gives $\Pi = diag(0_m\ \Pi_o)$. ❐

Proposition B.2. A balanced representation of a system with integrators (A_b, B_b, C_b, D) is obtained by the transformation T_b

$$A_b = T_b^{-1} A T_b, \quad B_b = T_b^{-1} B, \quad C_b = C T_b \tag{B-6}$$

where

$$T_b = T_1 T_2 \tag{B-7}$$

The transformation T_1 turns A into block-diagonal form $A_1 = diag(0_m\ A_o)$, e.g., into real modal form

$$A_1 = T_1^{-1} A T_1 \tag{B-8}$$

and the transformation T_2 is represented by the form $T_2 = diag\ (I_m\ T_{bo})$, where I_m is an identity matrix of order m, and T_{bo} balances A_o

$$A_{bo} = T_{bo}^{-1} A_o T_{bo} \tag{B-9}$$

Proof. Immediate, by introducing Eq. (B-5) into Eq. (B-3a). ❐

The antigrammian Π of a balanced system with integrators is ordered increasingly

$$\Pi = diag(\pi_1, \pi_2, \ldots, \pi_{n-1}, \pi_n) \tag{B-10}$$

where $\pi_i \geq 0$ and $\pi_{i+1} \geq \pi_i$ for $i = 1, \ldots, n$, with the first m singular values at zero, $\pi_i = 0$ for $i = 1, \ldots, m$. The system is reduced by truncating the last $n-k$ states of the balanced representation and leaving its first k states. Let the matrices A_b, B_b, and C_b be partitioned conformably

$$A_b = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B_b = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \quad C_b = [C_1 C_2] \tag{B-11}$$

then the reduced system representation (A_r, B_r, C_r, D_r) is $A_r = A_{11}, B_r = B_1, C_r = C_1$, and D remains unchanged. The system can also be reduced by the enhanced low-frequency approximation, see [13,14]

$$\left.\begin{aligned} A_r &= A_{11} - A_{12} A_{22}^{-1} A_{21}, \\ B_r &= B_1 - A_{12} A_{22}^{-1} B_2, \\ C_r &= C_1 - C_2 A_{22}^{-1} A_{21}, \\ D_r &= D - C_2 A_{22}^{-1} B_2 \end{aligned}\right\} \tag{B-12}$$

A Portable Ka-Band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part I. Design and Ground Tests

T. Y. Otoshi, S. R. Stewart, and M. M. Franco
Ground Antennas and Facilities Engineering Section

As described in previous articles, a unique experimental method was used to test the new Beam Waveguide (BWG) antenna at Deep Space Station 13 in the Goldstone Deep Space Communications Complex near Barstow, California. The methodology involved the use of portable test packages to make measurements of operating noise temperatures and antenna efficiencies (as functions of antenna pointing angles) at the Cassegrain focal point and the final focal point located in a subterranean pedestal room. Degradations caused by the BWG mirror systems were determined by making comparisons of the measured parameters at the two focal points of the antenna. Previous articles were concerned with the design, performance characteristics, and test results obtained with an X-band test package operating at 8.45 GHz. This article discusses a Ka-band test package designed for operation at 32 GHz. Noise temperature measurement results are presented for the Ka-band test package in an "on-the-ground" test configuration.

I. Introduction

Previous articles [1,2] have given detailed descriptions of the test package design, performance characteristics, and test results for the X-band package in "on-the-ground" and "on-the-antenna" test configurations at 8.45 GHz. It was shown that the X-band test package was used successfully to test the new DSS 13 BWG antenna at 8.45 GHz.

This article is the third in a series of reports on the measured performance of the new BWG antenna. In this article, the design of a test package developed for Ka-band is described. In addition, measured operating noise temperatures for the Ka-band test package in the on-the-ground test configuration are presented. A companion article, "Part II," in this issue presents the results of tests performed on the new BWG antenna with the Ka-band test package.

II. Ka-Band Test Package Design

Figure 1 shows the system block diagram of the Ka-band test package. Depicted are typical Cassegrain front-end microwave components, such as a horn, polarizer, round-to-rectangular transition, and a waveguide switch. When the Ka-band WR28 switch is in the antenna path

configuration, the switch connects the horn assembly to a WR28 coupler followed by a cryogenically cooled high-electron-mobility transistor (HEMT) and a downconverter. The insertion loss of the path between the horn aperture and the input of the HEMT at 32 GHz is 0.27 dB ±0.02 dB.

When the switch is in the ambient-load path configuration, the switch connects an ambient thermal reference load to the HEMT. System linearity calibrations are performed with a remotely controlled noise diode assembly. Noise temperature calibrations are performed with an ambient-load thermal reference termination, in which is embedded a digital readout thermometer. For this Ka-band test package, the microwave signal is downconverted to an intermediate frequency (IF) of 60 MHz and sent via a coaxial cable to a power-meter total-power radiometer system.

The requirement of testing the antenna at F1 and F3 required the test package to be convertible from a 29- to 23-dBi horn configuration. The conversion is accomplished through the use of horn extensions of the same taper going from an aperture diameter of about 5.122 in. to 1.986 in. Figures 2 and 3, respectively, are photographs of the fabricated and assembled Ka-band test package in its 29-dBi horn configuration for testing the system on the ground. The test package in the 29-dBi horn configuration, as measured from the bottom of the lower frame to the horn aperture, is about 8.5 ft high, while the 23-dBi configuration is about 14.3 in. shorter.

The Ka-band HEMT assembly shown in Fig. 2 is a four-stage amplifier with an overall gain of 24.4 dB and an effective input noise temperature of 56.6 K. Following the HEMT assembly are waveguide components and a Ka-band downconverter assembly (Fig. 3). A block diagram of the Ka-band downconverter assembly is given in Fig. 4. The vital components of this downconverter assembly are a stainless steel WR28 guide section for thermal isolation, a 32-GHz bandpass filter, a 3.5-dB noise figure mixer, an IF bandpass filter, and a thermoelectric temperature controller. The system operates double sideband, where input signals in the bands (31.89 to 31.99-GHz) and (32.01 to 32.11 GHz) become downconverted to an IF band of 10 to 110 MHz, whose center frequency is 60 MHz.

Early in the development stage, considerable gain instability problems were experienced with the Ka-band downconverter. The problems were finally overcome with the installation of the thermoelectric temperature controller (see Fig. 3), which maintained the temperature inside the downconverter to within ±0.4 deg C. After installation of the thermoelectric cooler, the gain change during a typical day was reduced from 0.95 dB to about 0.13 dB.

The Ka-band noise diode assembly (Fig. 5) provides an injection signal of about 81 K into the HEMT and is used to check system linearity. Measurements of the excess noise injected are made when the waveguide switch is first in the antenna position and then in the ambient load position. The difference in the excess noise measured in the two waveguide positions provides a means for determining noise-temperature measurement-system nonlinearity.

III. Radiometer System

Although the radiometer computer system is not part of the test package, the test package had to be designed specifically to interface with the computer-controlled radiometer system. For either the X- or Ka-band test packages, the downconverted output is sent to the computer-controlled radiometer system via a coaxial cable.

Figure 6 is a block diagram showing the interface between the Ka-band test package and the total-power radiometer system. The IF noise power is measured with a Hewlett-Packard 436A Power Meter. An IBM personal computer sends commands to change the waveguide switch to a thermal reference load position and reads the digital information on the thermal reference load to a resolution of ±0.01 deg C. Power meter readings at the zero and ambient load positions provide a calibration curve of power meter readings versus operating noise temperature. The computer commands the noise injecting diode to turn on and off in the ambient load and antenna positions, and performs calculations of system linearity, gain changes, and operating noise temperature. Calibrations are performed every 30 minutes or as often as commanded, and operating noise temperatures are computed every second or integrated over a desired time interval and then displayed on the screen and at the line printer. The data are automatically stored on the computer disk file.

For good operation for this particular Ka-band test package with respect to power meter resolution and system linearity, it was found that the input to the power meter sensor should be about 700 nanowatts when the waveguide switch was in the antenna path configuration. The adjustment of power levels was achieved with appropriate padding in the system and use of a variable IF attenuator in front of the power meter sensor.

Calibration and error analyses equations for the calibration method of this measurement system are based on

equations given in [3]. An option is available in the computer data reduction program to correct the measured operating noise-temperature values for nonlinearity [4]. More discussions of the calibration method and associated errors are given in Appendix A.

IV. Test Results

Figures 7(a) and 7(b) show examples of the overall system performance of the Ka-band test package on the ground at DSS 13 when the thermoelectric temperature controller was in operation. The symbols show data points for mini-calibrations performed every 30 min. During the 25-hr test period shown, the linearity factor remained nearly constant between 0.99 and 0.98. The gain factor varied between 1.00 and 0.99, which corresponds to a peak-to-peak gain change of only 0.04 dB. As may be seen in Fig. 7(b), the ambient-load physical temperature varied from a minimum of 2.7 to 16.1 deg C. Comparisons of the ambient-load physical temperature and gain factor plots show that they are correlated. Tests have shown that for ground tests, the outside air temperature and ambient-load physical temperatures track within a few deg C.

Noise-temperature symbols are used in the following equations, as well as in a table of results. For the reader's convenience, the symbols are defined in Table 1. When the test package is on the ground, the general expression for the operating noise temperature is

$$T_{op} = \frac{T'_{cb}}{L_{atm} L_{wg}} + \frac{T_{atm}}{L_{wg}} + T_{wg} + T_{hemt} + T_{fup} \qquad (1)$$

Under standard conditions at 32.0 GHz, the component values are $T'_{cb} = 2.0$ K, $T_{atm} = 7.02$ K, $T_{wg} = 17.67$ K, $T_{hemt} = 56.6$ K, $T_{fup} = 1.8$ K, $L_{atm} = 1.02683$ (corresponding to 0.1150 dB), and $L_{wg} = 1.06414$ (corresponding to 0.27 dB). Substitutions of the values into Eq. (1) result in a predicted T_{op} of 84.5 K.

Table 2 shows a tabulation of the Ka-band measurements for ground observations from October 12, 1990 through January 31, 1991. The grand average of the T_{op} values shown in Table 2 is 84.7 K, with a peak-to-peak deviation of $+1.6/-1.7$ K. The uncertainty associated with the grand average is estimated to be of 1.5 K (one standard deviation). This measured value of 84.7 K is in close agreement with the predicted value of 84.5 K.

Included in Table 2 are tabulations of the observation periods, the measured operating noise-temperature values, weather information, and normalized values after corrections were made for weather and the waveguide physical temperature values. This table shows the methodology used for normalizing zenith operating noise-temperature data to standard conditions. The normalized value of T_{op} is computed from

$$T_{op,n} = T_{op} + \frac{T'_{cb}}{L_{wg}} \left(\frac{1}{L_{atm,s}} - \frac{1}{L_{atm}} \right) + \frac{1}{L_{wg}} (T_{atm,s} - T_{atm}) + (T_{wg,s} - T_{wg}) \qquad (2)$$

where T_{op}, T_{atm}, T_{wg}, and L_{atm} are, respectively, the average measured T_{op}, computed T_{atm}, T_{wg}, and L_{atm} values given in Table 2. The values used for normalizing the measured values were $T_{atm,s} = 7.02$ K (for a 32-GHz Goldstone average clear atmosphere) and $L_{atm,s} = 1.02683$ (corresponding to 0.1150 dB). Other Ka-band (32.0 GHz) values used for normalizations were $T'_{cb} = 2.0$ K, $L_{wg} = 1.06414$ (corresponding to 0.27 dB), $T_{wg,s} = 17.67$ K for the above L_{wg}, and a standard physical waveguide temperature of 20 deg C.

At Ka-band, weather changes can cause significant variations in atmospheric noise contributions to operating noise temperatures. As may be seen in Table 2, the applications of weather corrections enabled significant improvements to be made to the operating noise temperatures for purposes of comparisons for measurements made at time periods separated by days and months.

Figure 8 shows curves of Ka-band atmospheric noise temperatures as functions of typical weather parameters during a calendar year at DSS 13. It is of interest to note that at Ka-band, typical weather conditions during the year at DSS 13 can cause the atmospheric noise temperatures to vary from 4.5 to 31.2, which represents a peak-to-peak variation of about 27 K. Method 1 of computer program SDSATM7M.BAS[1], described in [2], was the method used to derive values for Fig. 8. Method 1 was also used to make weather corrections for the noise temperature results presented in this article.

V. Concluding Remarks

Numerous Ka-band total-power radiometer tests on the ground showed that with a thermoelectric temperature

[1] Courtesy of S. D. Slobin of the Jet Propulsion Laboratory, Pasadena, California. SDSATM7M.BAS is a modified version of SDSATM4.BAS, but gives the same answers.

controller for the downconverter, gain changes were held to less than 0.2 dB over a 24-hour period, which is about as good as can be expected with a Ka-band total-power radiometer. Even when gain changes were large (before installation of the thermoelectric temperature controller), some of the gain instability problems were overcome by performing mini-calibrations more frequently than the normal 0.5-hr interval and by using the mini-calibration data to compute near-real-time corrections for gain changes. The combination of using a radiometer with small gain changes and performing more frequent mini-calibrations enabled better quality noise-temperature data to be obtained than was originally thought possible from a Ka-band total-power radiometer system.

Acknowledgments

The successful assembly and operation of the Ka-band test package are due to many people. D. Neff provided needed Ka-band parts and test equipment during critical phases of Ka-band test package testing. Although many difficulties were initially experienced with the Ka-band HEMT, these difficulties were overcome through the efforts of G. Ortiz, J. Loreman, L. Tanida, and S. Petty. The Ka-band downconverter was designed and fabricated by C. Risch and B. Finamore. Preliminary work on the Ka-band noise diode assembly was done by R. Strickland. All the above named are from the Radio Frequency and Microwave Subsystems Section.

The power supply and thermoelectric cooler for the downconverter were incorporated by R. Denning and R. Wetzel of the Microwave Observational Systems Section. Group Supervisor N. Yamane provided the technical and administrative support needed to get the thermoelectric assembly work done on a crash program basis at a critical phase of this Ka-band project. The thermoelectric cooler was the critical element that was needed to make the downconverter gain stable. C. T. Stelzried, of the TDA Technology Development Office, provided valuable consulting assistance on the subject of noise temperatures and radiometer calibrations.

Table 1. Definitions of symbols and abbreviations

Symbol	Definition
T_{cb}'	Effective noise temperature contribution to T_{op} from the cosmic background radiation, K. This value is a function of frequency and will differ from the actual cosmic background noise temperature of 2.7 K [5][a].
T_{cb}	Cosmic background radiation noise temperature, nominally 2.7 K
h	Planck constant
f	Frequency, Hz
k	Boltzmann constant
T_{atm}	Atmospheric noise temperature, K
T_{wg}	Noise temperature due to waveguide loss between the horn aperture and the input flange of the HEMT, K
T_{hemt}	Effective noise temperature of the HEMT as defined at the input flange of the HEMT, K
T_{fup}	Effective noise temperature of the follow-up receiver (downconverter + cables + power meter, etc.) as defined at the input flange of the HEMT, K
T_{op}	Operating noise temperature as defined at the input flange of the HEMT, K
RH	Relative humidity
HEMT	High-electron-mobility transistor
L_{wg}	Loss factor for the waveguide between the horn aperture and the input flange of the HEMT, power ratio > 1
L_{atm}	Loss factor of the atmosphere, power ratio > 1

[a] $T_{cb}' = T_{cb}\left[\frac{x}{\exp(x)-1}\right]$ where $x = \frac{hf}{kT_{cb}}$. (See details in [5].)

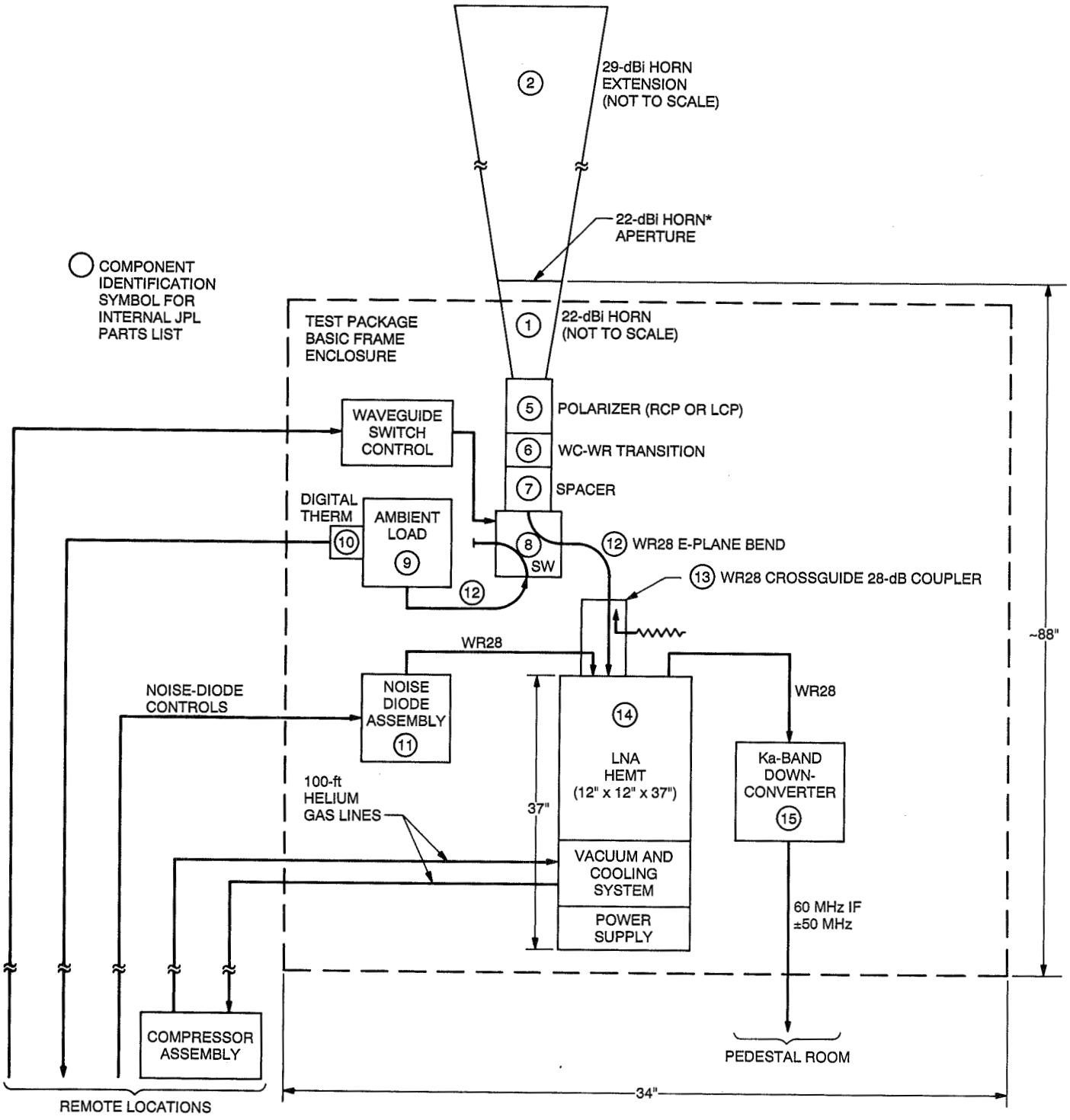
Table 2. Ka-band (32-GHz) measured zenith operating noise temperatures corrected for weather and waveguide noise changes. Ka-band test package is on the ground.

Observation	Average Measured T_{op}, K	Average Weather During Obsv.	Computed T_{atm}, K	Computed L_{atm}, ratio	Physical Waveguide Temp., deg C	T_{wg}, K	Normalized[a] T_{op}, K
10/12/90 DOY 285 0900–1250 UT	85.5	896.3 mb 17.9 deg C 15.8% RH	6.83	1.0258 (0.1104 dB)	18.9	17.61	85.7
11/09/90 DOY 313 1500 UT	87.5[b]	903.6 mb 14.9 deg C 31.7% RH	8.58	1.0324 (0.1386 dB)	15.0	17.38	86.3
01/19/91 DOY 019 1000–1500 UT	83.9	895.8 mb 7.12 deg C 44.3% RH	7.91	1.0303 (0.1296 dB)	6.7	16.88	83.9
01/31/91 DOY 031 0300–1000 UT	80.5	903.1 mb 5.3 deg C 13.7% RH	5.50	1.0211 (0.0906 dB)	1.3	16.55	83.0

Grand average = 84.7
Peak deviations = +1.6/−1.7

[a] Normalized T_{op} values were computed through the use of Eq. (2).

[b] The Ka-band waveguide was cleaned and realigned in the test package.



*ADDED A 0.375" LENGTH HORN EXTENSION AT THIS REFERENCE PLANE TO CONVERT FROM A 22-dBi TO 23-dBi HORN CONFIGURATION

Fig. 1. Ka-band test package system.

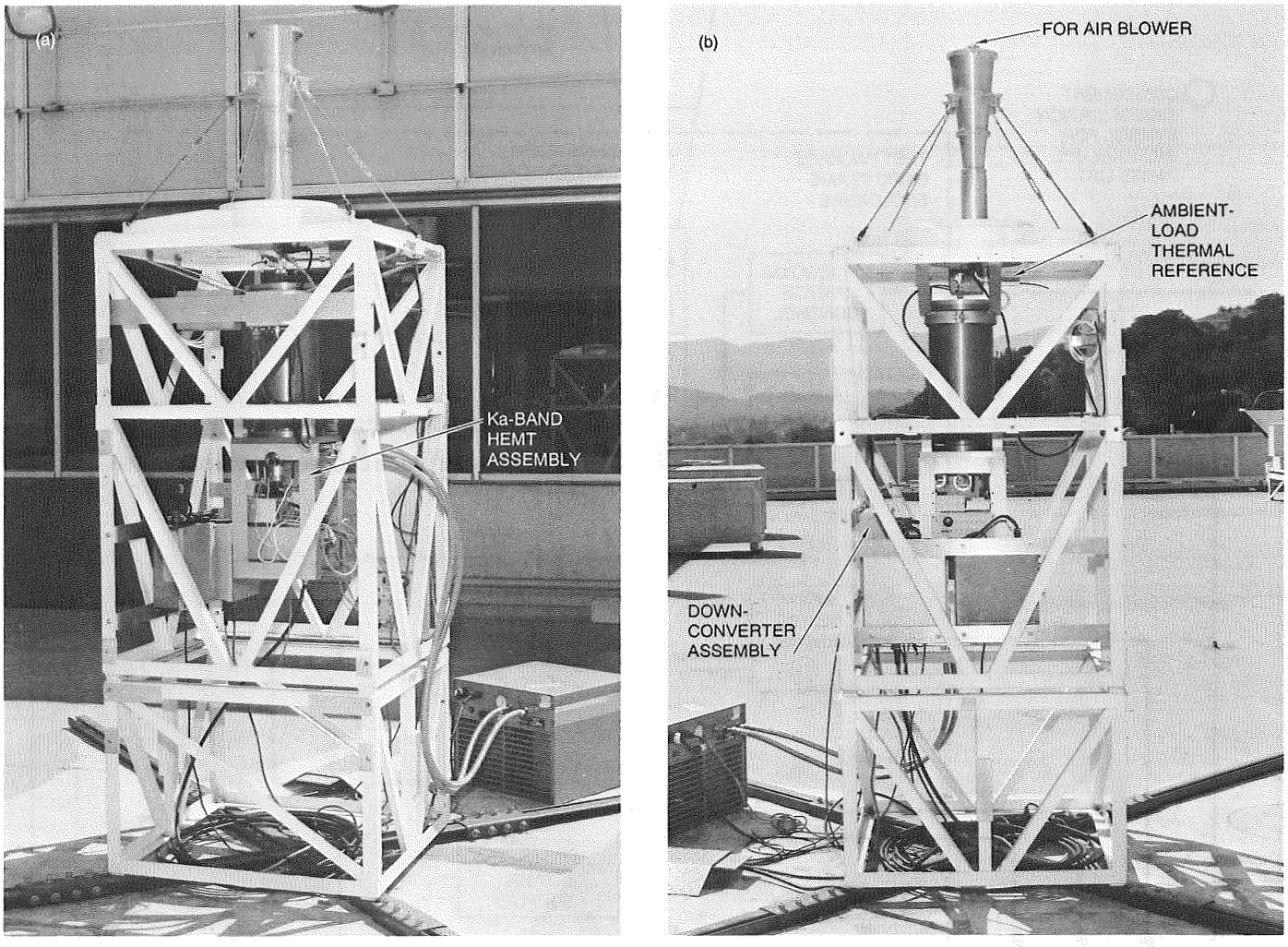


Fig. 2. Ka-band test package in its 29-dBi horn configuration for testing the system on the roof of the Telecommunications Building: (a) showing Ka-band HEMT and (b) showing Ka-band ambient-load thermal reference and downconverter.

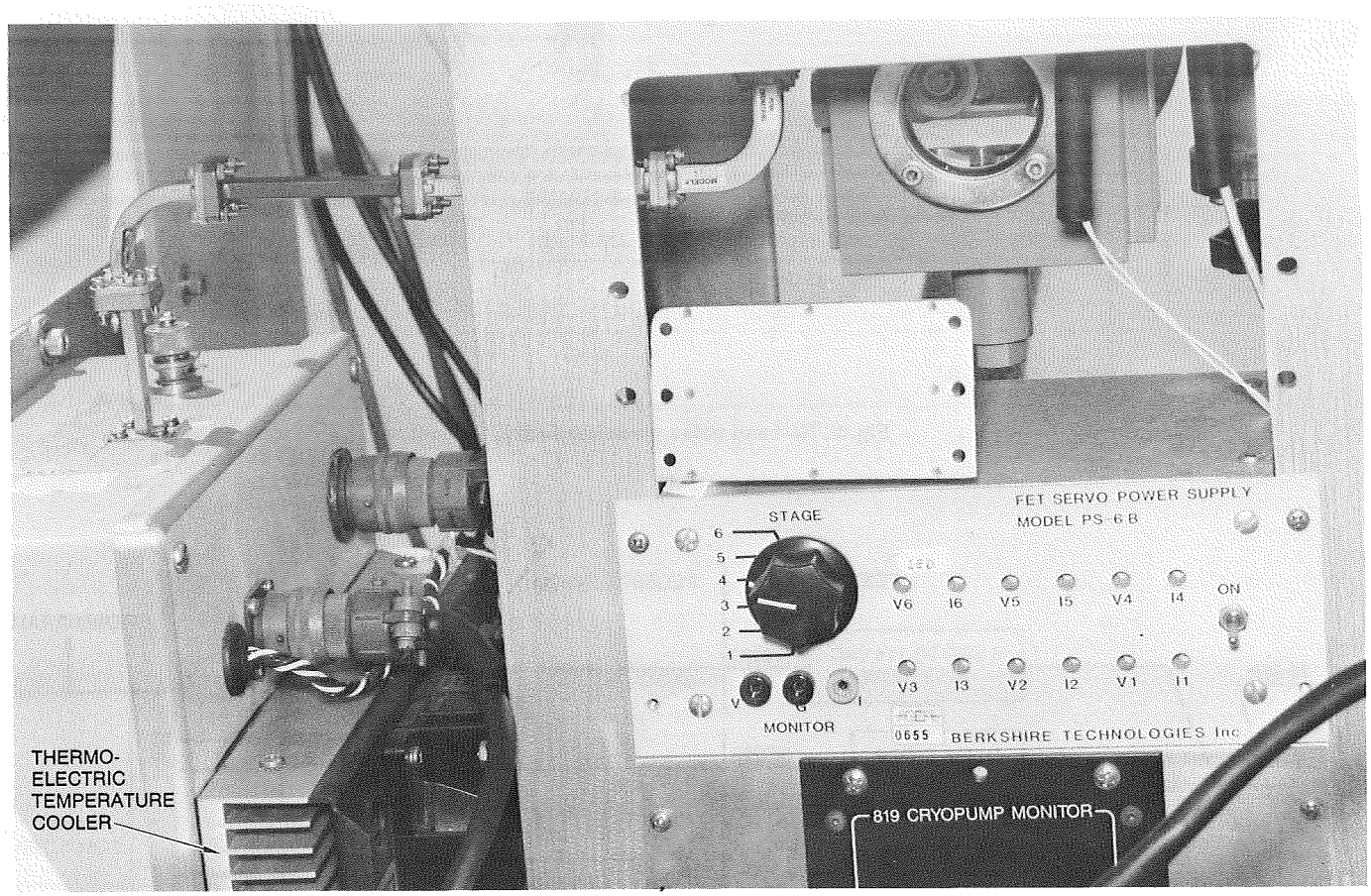


Fig. 3. Partial view of the post-HEMT assembly, including waveguide, downconverter, and thermal electric cooler assemblies.

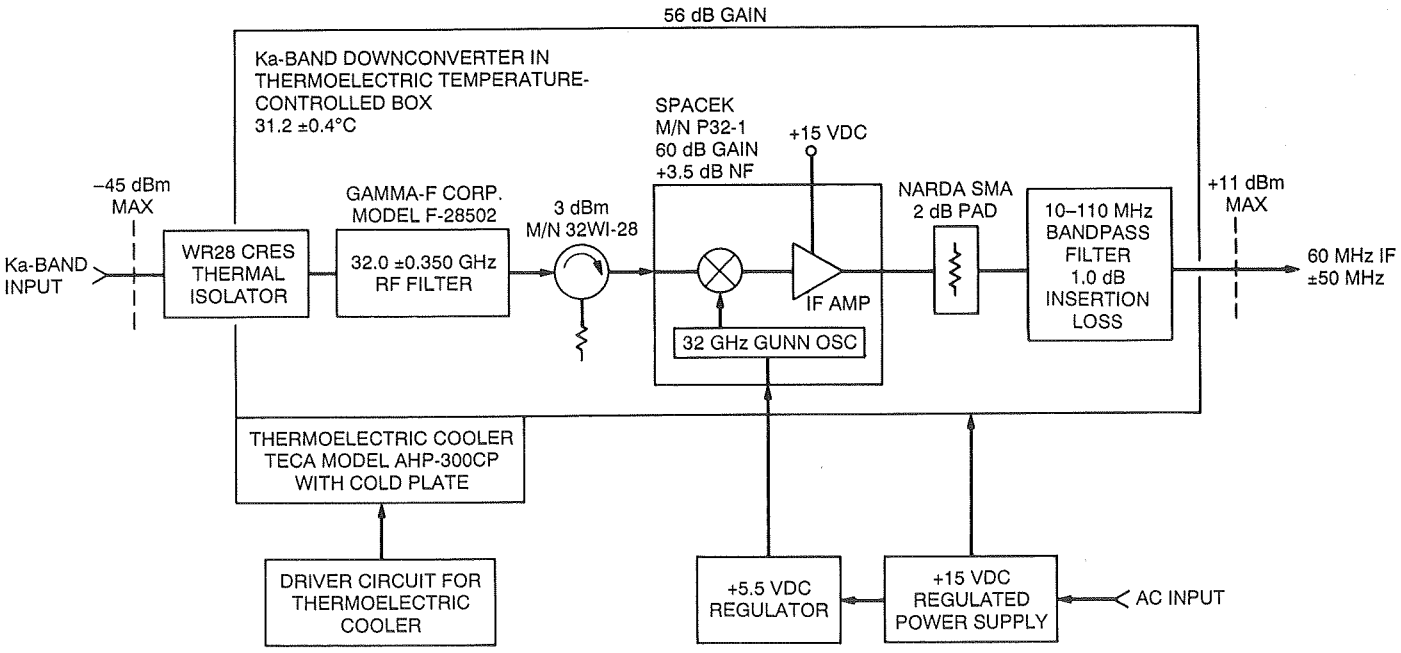


Fig. 4. Ka-band downconverter assembly.

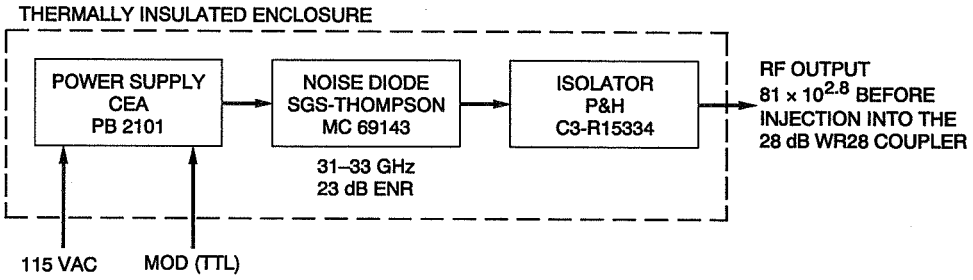


Fig. 5. Ka-band noise diode assembly.

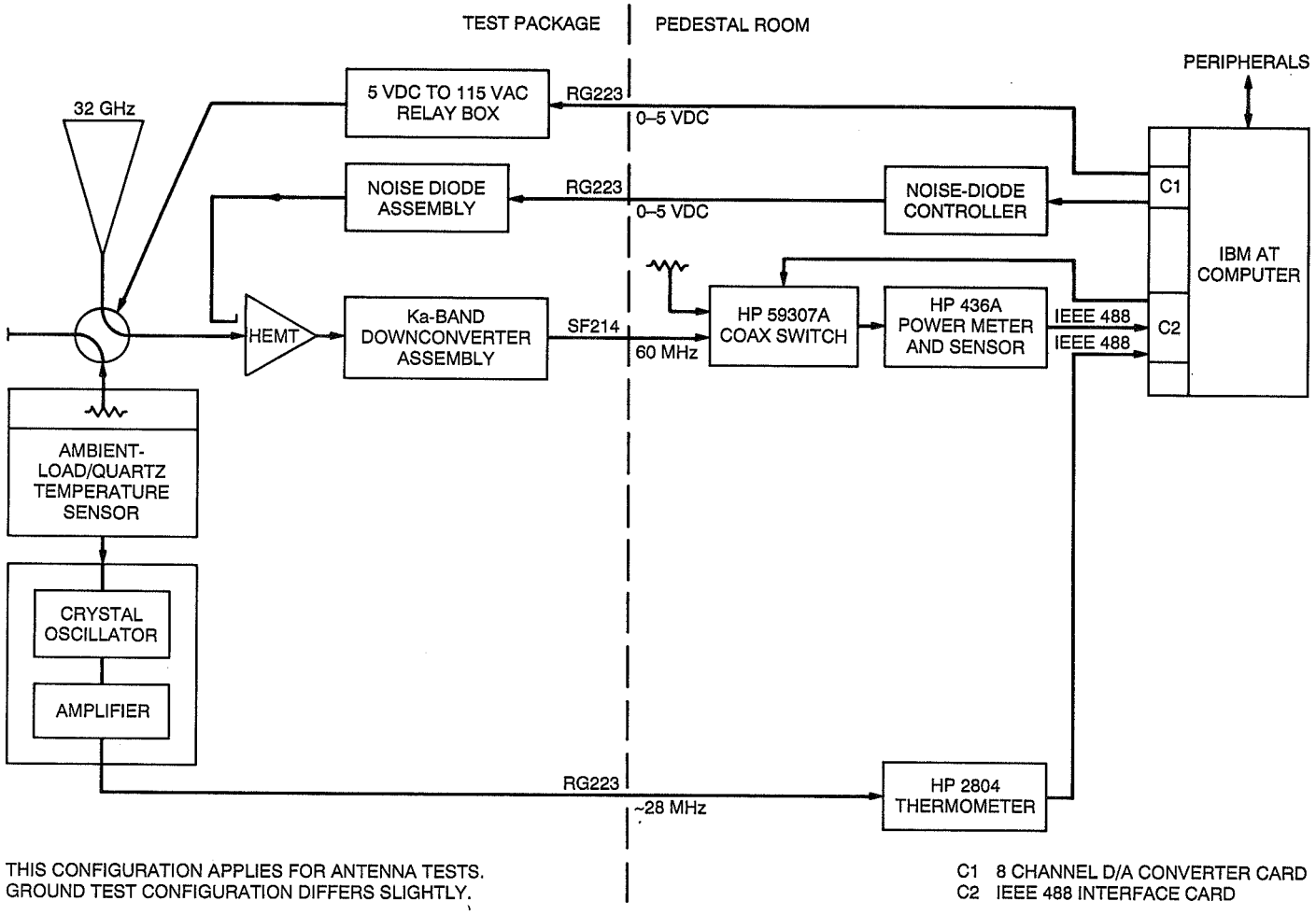


Fig. 6. The interface between the Ka-band test package and the total-power radiometer system.

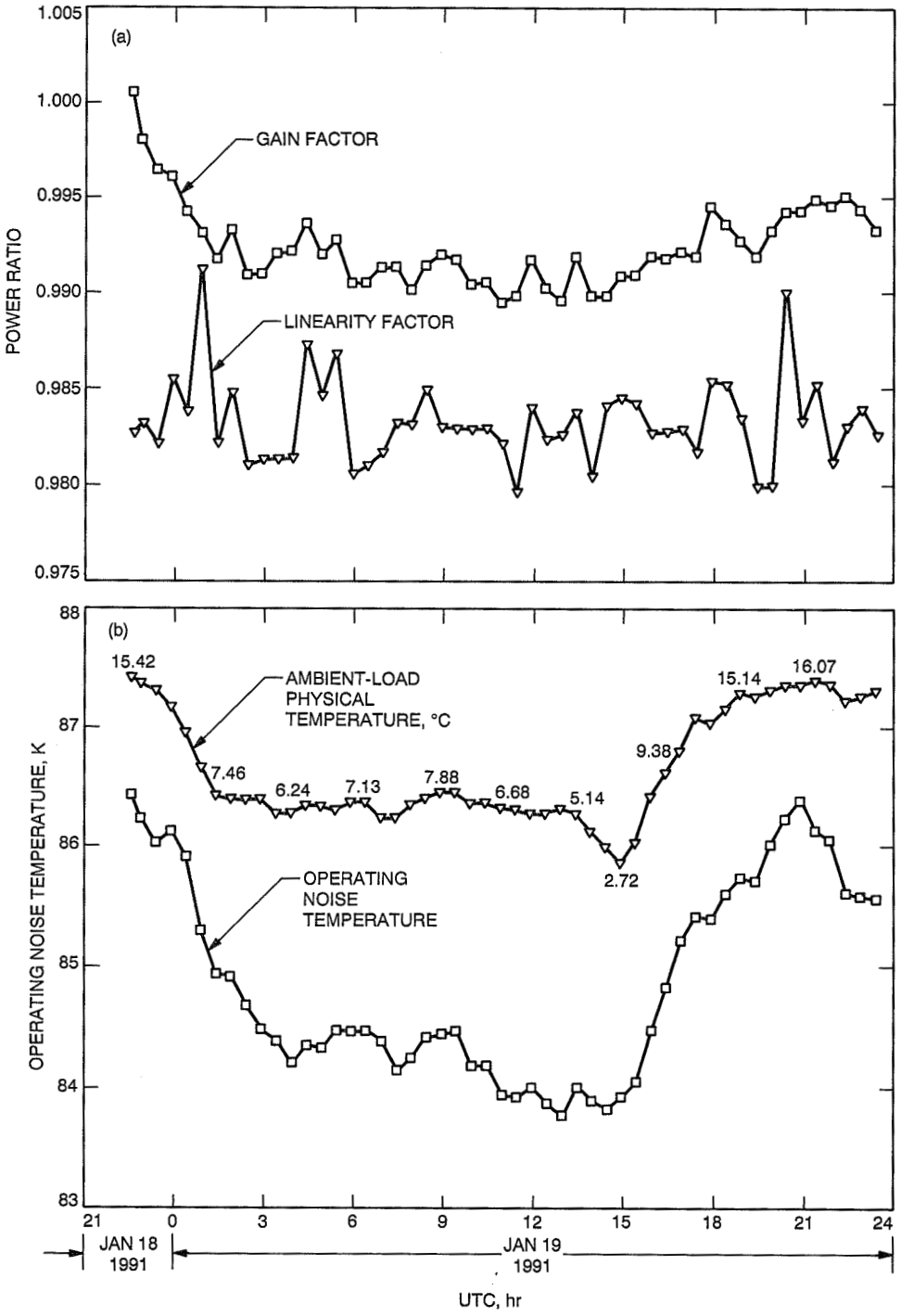


Fig. 7. Mini-calibration data during January 18 and 19, 1991. The Ka-band test package is outside of Building G60 (the Administrative and Laboratory building) at DSS 13: (a) gain and linearity factors and (b) operating noise and ambient temperatures.

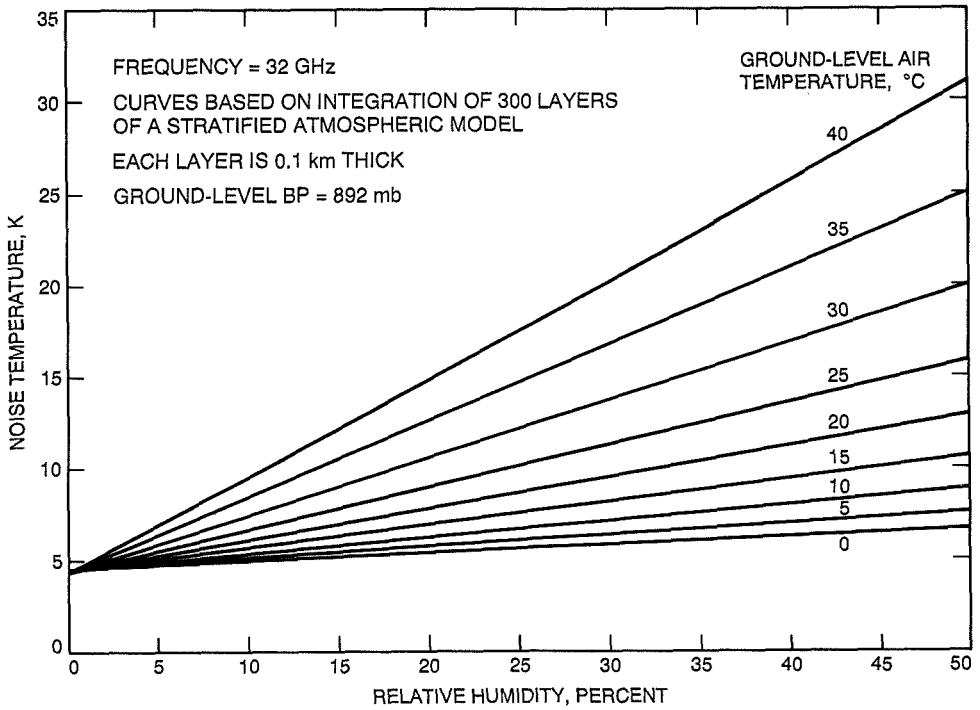


Fig. 8. Ka-band atmospheric noise-temperature contributions versus relative humidity and ground-level air temperature at DSS 13. (Courtesy of S. Slobin, Jet Propulsion Laboratory, Pasadena, California.)

Appendix A

Calibration Method and Error Analysis

This Appendix is written in response to numerous requests to describe the calibration system in more detail, and also to provide calibration and error analysis equations applicable to the test package noise-temperature measurement system. The measurement method and total-power radiometer equations follow those originally developed by Stelzried [3].

The principle of operation of the calibration system is illustrated in Fig. A-1. Two power meter readings are required to establish a linear power meter reading versus operating noise temperature calibration curve. The first required point of the curve corresponds to the power meter reading and an effectively zero[1,2] operating noise temperature when the power meter is zeroed by means of a remotely controllable coaxial switch that is terminated with a coaxial termination (Fig. 6). The second required point on the calibration curve corresponds to the power meter reading and the operating noise temperature when the waveguide switch is operated to be in the "ambient-load" path position (configuration). In this switch position, the waveguide ambient load becomes the input termination of the HEMT. The ambient load for this calibration system is used as the primary thermal noise reference standard. Based on the assumption that the calibration curve is linear, the two points are used to determine the constants (slope and intercept) of the calibration curve.

When the measurement system is used for measuring antenna operating noise temperatures, the waveguide switch is operated to be in the antenna path position. The power meter value is read by the computer and then the operating noise temperature (on the antenna) is computed from the power meter reading and the constants of the linear calibration curve.

It is assumed that the operating noise temperature in the ambient load configuration is well below the saturation point on the curve shown in Fig. A-1. A linearity check is accomplished by turning the noise diode on and off when the waveguide switch is first operated to be in the antenna path position and then to the ambient load path position. The excess noise due to the noise diode is computed and the difference of excess noise measured in the two switch positions provides a measure of the system's nonlinearity. An option in the computer program is provided for making corrections for nonlinearity through the use of equations derived by Stelzried in [4].

For purposes of performing an error analysis of the measured value of the operating noise temperature (on the antenna), the following calibration equations are applicable. It should be pointed out that in the main body of this article, T_{op} was used to denote the operating noise temperature in the antenna configuration. In the analysis below, T_{opa} and T_{oph} are used to denote operating noise temperatures for the antenna and ambient-load system configurations. The following general equations are applicable to the X- or Ka-band test package noise-temperature measurement systems:

$$T_{opa} = T_{oph}\frac{P_{opa}}{P_{oph}} \tag{A-1}$$

$$T_{oph} = T_{pc} + 273.16 + T_e \tag{A-2}$$

$$T_e = T_{hemt} + T_{fu} \tag{A-3}$$

where many of the symbols have already been defined elsewhere in this article. Definitions for symbols not previously defined are T_{pc} = ambient load physical temperature in deg C, P_{oph} = power meter reading in nW when the waveguide switch is in the ambient-load path position, and P_{opa} = power meter reading in nW when the waveguide switch is in the antenna path position.

For this particular calibration technique to be useful in practice, it is required that T_e be known accurately at the HEMT input flange, where T_e is defined above by Eq. (A-3). For T_e to be known accurately, it is necessary for T_{hemt} to be measured extensively in the laboratory prior to the installation of the HEMT assembly into the test package. Furthermore, it is assumed for this calibration method that T_{hemt} does not change with time or with test package motions or antenna motions. It has been found, based on two or more years of operation in the field, that T_{hemt} does not change significantly with time or antenna motion and the above assumptions are valid ones. Hence, the cooled HEMT is an ideal low-noise amplifier to

[1] Typically, this reading is less than 3×10^{-4} K for a system having an overall gain of 60 dB. The exact value can be calculated from an equation given by Stelzried[2].

[2] C. T. Stelzried, "DSS 13 Radiometer System Status and Performance," Interoffice Memorandum CTS-90-001 (internal document), Jet Propulsion Laboratory, Pasadena, California, January 30, 1990.

use for the noise measurement technique employed for the test package calibration system. The follow-up receiver noise temperature contribution (as defined at the HEMT input) is usually measured after installations through the use of a Y_{on-off} method. This follow-up noise temperature contribution is also required to be constant with time. If the HEMT gain is sufficiently high (> 30 dB) and the follow-up receiver effective input noise temperature is less than 500 K, the follow-up receiver contribution will be less than 0.5 K. Therefore, there is less probability that instabilities in the operating noise temperature measurements will come from the follow-up contribution.

Another requirement for the calibration system is that the ambient-load physical temperature T_{pc} can be measured to a high degree of accuracy (better than 0.2 deg C). If the requirements for both T_e and T_{pc} are met, then updated measurements of T_{pc} enable updated values of T_{oph} to be calculated and enable gain changes in the radiometer system to be isolated. Then, corrections can be determined for *gain changes* and applied to raw (uncorrected) measured operating noise temperature values to obtain what the computer program has labeled as "corrected system temperatures." Unfortunately, this terminology has caused some confusion in data comparison because Stelzried [4] also uses the same term to mean an operating noise temperature corrected for both *gain changes* and *nonlinearity*. The process of obtaining updated values of T_{pc} and corrected T_{op} is accomplished automatically at any desired periodic interval or instantaneously upon command by an experimenter through the use of a single computer keystroke operation. The procedure for obtaining data for an updated noise temperature calibration curve (Fig. A-1), making corrections for gain (and linearity, if desired), is referred to as a "mini-cal" (mini-calibration) operation[2].

If T_e is not known accurately or changes with time, the errors produced on the measurement of T_{op} can be calculated from the equations given below. Errors caused by other error sources, such as uncertainties of the ambient-load physical temperatures and power-meter reading errors can be calculated from the error analysis equations presented below.

Following the method given in [3], the root sum square (rss) of the individual error contributions to the overall T_{opa} measurement error is

$$E_{Topa} = \mathrm{SQRT}(E^2_{Topa/Tpc} + E^2_{Topa/Te} + E^2_{Topa/Popa} + E^2_{Topa/Poph} + E^2_{Topa/mm}) \quad \text{(A-4)}$$

where SQRT denotes square root and the general symbol of the form $E_{y/x}$ is used to denote the error in the value of y due to an error in x.

Then,

$$E_{Topa/Tpc} = \left|\frac{\partial T_{opa}}{\partial T_{pc}} E_{Tpc}\right| = \frac{T_{opa}}{T_{oph}} E_{Tpc} \quad \text{(A-5)}$$

$$E_{Topa/Te} = \left|\frac{\partial T_{opa}}{\partial T_{e}} E_{Te}\right| = \frac{T_{opa}}{T_{oph}} E_{Te} \quad \text{(A-6)}$$

$$E_{Topa/Popa} = \left|\frac{\partial T_{opa}}{\partial P_{opa}} E_{Popa}\right| = \frac{T_{oph}}{P_{oph}} E_{Popa} \quad \text{(A-7)}$$

$$E_{Topa/Poph} = \left|\frac{\partial T_{opa}}{\partial P_{oph}} E_{Poph}\right| = \frac{T_{opa}}{P_{oph}} E_{Poph} \quad \text{(A-8)}$$

Equations for $E_{Topa/mm}$ are involved and will not be presented in this Appendix. Computer programs for a personal computer have been written for the mismatch error equations given by Otoshi [6]. Sufficiently good estimates of the worst-case errors due to mismatch can be obtained through the use of simplified error formulas given in [3].

In the above, it can be seen that in order to minimize the error on the measurement of T_{opa}, it is desirable to make the ratio of T_{oph}/T_{opa} as large as possible. This goal is achieved in practice by selecting an accurate hot reference load noise temperature that is much higher than the antenna noise temperature to be measured and by using a receiver with a low value of T_e. For the case where the ambient reference termination provides the hot-source reference noise temperature, it is desirable that the ambient-load operating-noise temperature be at least 10 times higher than the antenna operating-noise temperature. Examples of X- and Ka-band test package error calculations are given in Table A-1.

Table A-1. Examples of errors in measurement of antenna operating noise temperature

	X-Band	Ka-Band
Measured or known parameter values		
T_{opa}	30 K	87 K
T_e	13.4 K	58.4 K
T_p	293.16 K	293.16 K
T_{oph}	306.56 K	351.56 K
P_{opa}	700 nW	700 nW
P_{oph}	7153 nW	2829 nW
1σ estimate of individual error sources		
E_{Tpc}	0.2 deg C	0.2 deg C
E_{Te}	0.2 K	0.5 K
E_{Popa}[a]	2.4 nW	2.4 nW
E_{Poph}[a]	15.3 nW	6.7 nW
Error source contributions		
$E_{Topa/Tpc}$	0.02 K	0.05 K
$E_{Topa/Te}$	0.02 K	0.12 K
$E_{Topa/Popa}$	0.10 K	0.30 K
$E_{Topa/Poph}$	0.06 K	0.21 K
T_{opa} rss error	0.12 K	0.39 K

[a] Assume that the power meter accuracy in nanowatts is given by $\pm(1 + 0.002 \times \text{PMR})$, where PMR is the power meter reading in nanowatts.

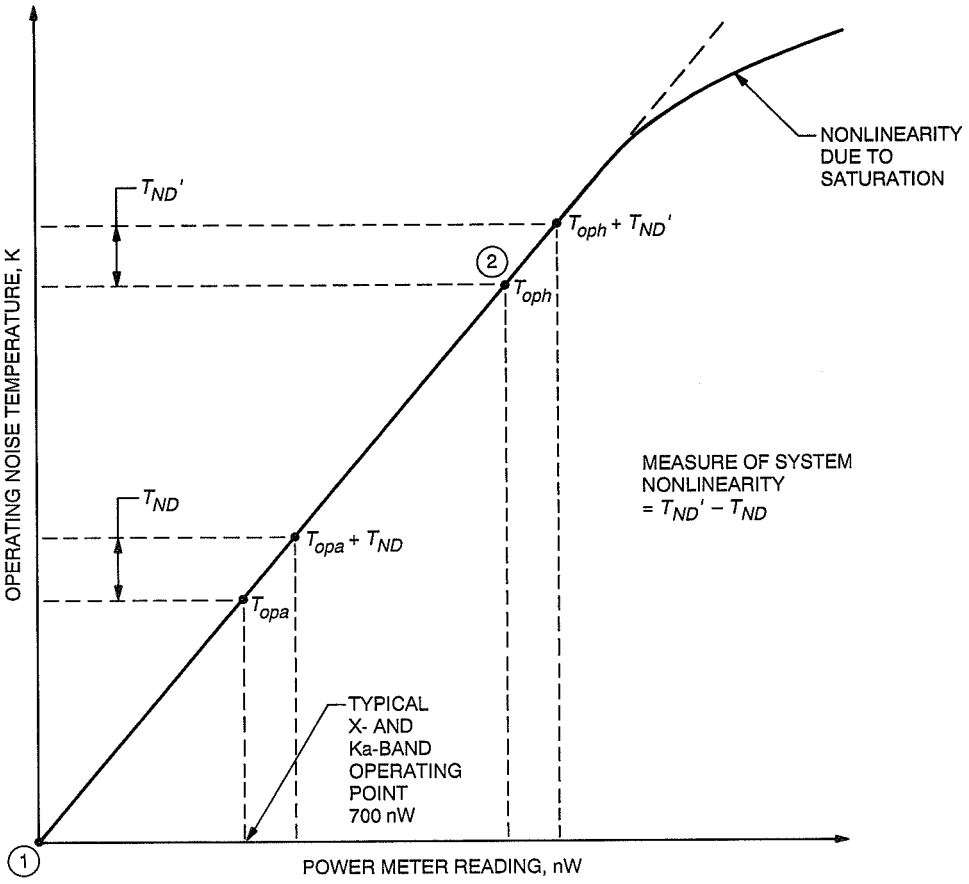


Fig. A-1. Example of a calibration curve for X- or Ka-band test package radiometer system.

References

[1] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable X-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part I: Design and Ground Tests," *TDA Progress Report 42-103*, vol. July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 135–150, November 15, 1990.

[2] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable X-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part II: Design and Ground Tests," *TDA Progress Report 42-105*, vol. January–March 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 54–68, May 15, 1991.

[3] C. T. Stelzried, "Operating Noise-Temperature Calibrations of Low-Noise Receiving Systems," *Microwave Journal*, vol. 14, no. 6, pp. 41–46 and 48, June 1971.

[4] C. T. Stelzried, "Non-linearity in Measurement Systems: Evaluation Method and Application to Microwave Radiometers," *TDA Progress Report 42-91*, vol. July–September 1987, pp. 57–61, November 15, 1987.

[5] C. T. Stelzried, "The Deep Space Network—Noise Temperature Concepts, Measurements, and Performance," JPL Publication 82-33, Jet Propulsion Laboratory, Pasadena, California, pp. 2-1–2-3, September 15, 1982.

[6] T. Y. Otoshi, "The Effect of Mismatched Components on Microwave Noise Temperature Calibrations," *IEEE Transactions on Microwave Theory and Techniques (Special Issue on Noise)*, vol. MTT-16, no. 9, pp. 675–686, September 1968.

C-4

TDA Progress Report 42-106

August 15, 1991

A Portable Ka-Band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part II: Tests on the Antenna

T. Y. Otoshi, S. R. Stewart, and M. M. Franco
Ground Antennas and Facilities Engineering Section

In "Part I" of this article, a description was given of a Ka-band test package developed to enable testing of the new DSS 13 34-m beam-waveguide (BWG) antenna at 32 GHz. Test results were given for the Ka-band test package in an "on-the-ground" test configuration. This article, "Part II," is a companion article concerned with Ka-band test results for the test package in an "on-the-antenna" test configuration. Included are Ka-band zenith noise-temperature values, tipping-curve data, and subreflector test results obtained at the Cassegrain focal point F1, as well as at the final BWG focal point F3 (located in a subterranean pedestal room). Test results show that, through the use of the Ka-band test package, the BWG antenna performance was successfully evaluated at Ka-band. The Ka-band test package operated well in all of the different antenna test configurations.

I. Introduction

As described in [1], a new 34-m beam-waveguide (BWG) antenna has been built at Deep Space Station 13 in the Deep Space Communications Complex at Goldstone, near Barstow, California. This new antenna is the first of the NASA tracking antennas to use a BWG design. A recent photograph of the newly constructed antenna is shown in Fig. 1. Figure 2 shows the various focal points (F1, F2, and F3) of the center-pass mode of this new BWG antenna. The experimental test program includes testing the new antenna at X-band (8.45 GHz), Ku-band (12 GHz), and Ka-band (32 GHz).

As described in previous reports [1,2], a unique methodology was used to test the new antenna at the different frequency bands. The methodology involved the use of portable test packages that could be transported to the various focal points (see Fig. 2). Measurements of operating noise temperatures and antenna efficiencies were made as functions of antenna elevation and azimuth angles at the different focal points of the BWG antenna. Comparisons of data measured at the different focal points enabled determinations of the performance degradations caused by various mirrors in the BWG system.

Previous reports [1,2] gave results of X-band test package measurements in "on-the-ground" and "on-the-antenna" test configurations. The results of Ka-band test package measurements in an on-the-ground test configuration were presented in a companion article, "Part I," in this issue [3].

This present article, "Part II," is the fourth of a series of reports on the results of measurements obtained through the use of portable microwave test packages to evaluate the performance of the new BWG antenna. Presented in this article are Ka-band zenith operating noise temperatures measured at focal points F1, F2, and F3, as well as Ka-band tipping curve and subreflector test data. Beam-waveguide antenna efficiencies measured with the X- and Ka-band test packages at F1 and F3 are presented in another article in this issue [4]. Another related article in this issue is concerned with the BWG antenna-pointing data obtained with the X- and Ka-band test packages [5].

II. Installations

Figure 3 shows the Ka-band 29-dBi horn test package installed on the antenna at the Casssegrain focal point F1. The structure shown supporting the Ka-band test package is a universal F1 mounting assembly that allows any of the three 29-dBi horn configuration test packages (X- ,Ku-, or Ka-band) to be interchanged and installed.

After completing measurements of noise temperature and antenna efficiencies at F1, the Ka-band test package was removed and then reconfigured to a 23-dBi horn configuration and installed at F3 (Fig. 4). The mounting table shown is a universal F3 mounting assembly that can support any of the test packages (X-, Ku-, or Ka-band) and provides three-axis adjustment of the test package location. Adjustments of ± 3 in. can be made along three orthogonal axes (i.e., the vertical direction, the radial direction towards and away from the BWG hub center, and the other transverse direction).

III. Test Results

A. Zenith Operating Noise-Temperature Measurements

Figure 5 shows plots of gain factor, linearity, and operating noise temperatures when the 29-dBi horn test package was mounted at F1. Figure 5(a) shows that over the 16-hour test period, the gain factor varied from 0.999 to 0.993, corresponding to a peak-to-peak gain change of only 0.03 dB. Figure 5(b) shows that during this same test period, the change in operating noise temperature was approximately 0.6 K peak-to-peak.

Figure 6 shows typical plots obtained when the test package was mounted at F3. Over a 14-hour period, the gain factor varied from 0.995 to 1.005, corresponding to a peak-to-peak gain change of about 0.04 dB. The change in operating noise temperature was approximately 0.8 K peak-to-peak.

It was fortuitous that Ka-band tests were done during winter months when the weather was cold and the humidity low. As may be seen in Figs. 5(b) and 6(b), the operating noise temperatures were reasonably constant for tests at both F1 and F3. It should be pointed out that in these plots, the ambient-load physical temperatures are correlated with the outside air temperatures at F1 (an outdoor environment), but uncorrelated when the test package is at F3 (an indoor environment). At Ka-band, operating noise-temperature values are very sensitive not only to outside air temperature and relative humidity, but also to variable cloud cover, atmospheric inversion layers, and uneven heating of the atmosphere.

For the operating noise-temperature values shown in Figs. 5(b) and 6(b), corrections were made for gain changes, but not for nonlinearity, which was typically less than 1 percent. Since the uncertainty of the linearity corrections are themselves about ± 1 percent, the small linearity corrections, if made, would not significantly enhance the quality of the operating noise-temperature data shown on these particular plots.

A summary of results of Ka-band zenith operating noise temperatures for the ground and F1, F2, and F3 locations is given in Table 1. The value shown in the table for each test configuration is the grand average of the average operating noise temperatures determined for the individual observing periods. The total span of time over which observations were made is about four months. Each grand average value shown in Table 1 is estimated to have an uncertainty of ± 1.5 K (one standard deviation).

Zenith operating noise temperatures for the F2 focal point were extrapolated from antenna tipping-curve tests and from a star track. Therefore, the confidence level of the data at F2 is not as high as at the other focal point locations where more data were obtained. Tests were done at F2 for diagnostic purposes to investigate a "hysteresis-like" characteristic on antenna efficiency measurements. The value obtained at F2 is included in Table 1 to enable interesting comparisons with values obtained at other focal points.

One of the primary goals of the experiment was to determine the amount of BWG antenna degradations caused by the mirror system. The methodology proposed to determine the degradations was to take the difference of values obtained at the various focal points. Table 2 shows differential noise temperatures for the various configurations. The "F1 – ground" configuration data have a value of 7.1 K due primarily to scattering from the subreflector and tripod, leakage through the main reflector surface,

and spillover from the main reflector. Another cause of noise temperature increase at F1 is a temporary perforated plate having 3/8-in.-diameter holes was used to cover the bypass shroud opening (Fig. 3) on the main reflector surface. Even though this plate is reflective at X-band, it begins to approach transparency at Ka-band frequencies. The ground noise contributions due to leakage through this perforated plate are, therefore, greater at 32 GHz than at 8.45 GHz. The amount of noise contribution might be estimated from information provided in [2] for the shroud opening (when uncovered) at 8.45 GHz. More noise might also have been contributed by numerous pieces of masking tape (Fig. 3) covering panel-adjustment bolt holes on the main reflector surface. The numerous pieces of masking tape, used to post panel-adjust information, were not removed after completing holography measurements at F1. Although the masking tape pieces might not have affected holography or X-band measurements, they might have increased reflector surface noise temperature contributions at Ka-band.

Comparisons of the "F3 – F1" values before and after mirror realignments show that realignment of the mirrors resulted in a lowering of the operating noise temperature by 3.8 K. The final "F3 – F1" value of 6.8 K provides a realistic measure of the degradation caused by six mirrors (including the ellipsoid) of the BWG system.

Table 3 is a worksheet showing how the final values summarized in Table 1 were derived. Shown are the observation periods, the measured operating noise-temperature values, weather information, and normalized values after corrections were made for weather and waveguide physical temperatures. In Table 3, the normalized T_{op} was computed from

$$T_{op,n} = T_{op} + \frac{T'_{cb}}{L_{wg}}\left(\frac{1}{L_{atm,s}} - \frac{1}{L_{atm}}\right) + \frac{1}{L_{wg}}(T_{atm,s} - T_{atm}) + (T_{wg,s} - T_{wg}) \quad (1)$$

where T_{op}, T_{atm}, T_{wg}, and L_{atm} are, respectively, the average measured T_{op}, computed T_{atm}, T_{wg}, and L_{atm} values given in Table 4, and where other noise-temperature symbols are defined in Table 1 of "Part I" [3]. The Ka-band (32.0 GHz) values used in Eq. (1) were T'_{cb} = 2.0 K [3], $T_{atm,s}$ = 7.02 K (for a Goldstone average clear atmosphere[1]), $L_{atm,s}$ = 1.02683 (corresponding to 0.1150 dB), L_{wg} = 1.06414 (corresponding to 0.27 dB), $T_{wg,s}$ = 17.67 K for the above L_{wg}, and a standard physical waveguide temperature of 20 deg C.

At Ka-band and higher microwave frequencies, weather changes can cause significant variations in the atmospheric noise contributions to the operating noise temperatures. In "Part I" [3], curves were presented showing that at Ka-band, typical weather conditions at DSS 13 during the year can cause the atmospheric noise temperatures to vary from 4.5 to 31.2 K, which represents a peak-to-peak variation of about 27 K. As may be seen in Table 3, the applications of weather corrections enabled significant improvements to be made in the comparisons of operating noise temperatures for the different test configurations—even when the measurements were performed on different days or months of different test configurations.

B. Cassegrain and BWG Antenna Loss Factors

It is of interest to determine equivalent loss factors for the BWG system. Stelzried and Otoshi [6] have shown that loss factors of very low-loss antenna waveguide components can sometimes be determined from radiometric methods when it is not possible to use conventional insertion-loss measurement techniques.

The derivations of the equations for BWG-antenna loss factors are straightforward, but lengthy, and will not be presented here. However, the basic equations needed to derive the loss factor equations will be presented. Let L_{F1} be defined as the loss factor associated with dissipative losses at the Cassegrain reflector surface, L_{F2} the loss factor of the BWG section between F1 and F2, and L_{F3} the loss factor of the BWG section between F1 and F3. The following basic equations apply for the measurements made with the test packages:

$$T_{sky} = \frac{T'_{cb}}{L_{atm}} + T_{atm} \quad (2)$$

$$T_{tp} = T_{wg} + T_{hemt} + T_{fu} \quad (3)$$

$$T_{op,g} = \frac{T_{sky}}{L_{wg}} + T_{tp} \quad (4)$$

$$T_{bwg,F1} = \left(1 - \frac{1}{L_{F1}}\right) T_p \quad (5)$$

$$T_{bwg,F2} = \left(1 - \frac{1}{L_{F2}}\right) T_p \quad (6)$$

[1] *Deep Space Network/Flight Project Interface Design Handbook*, Document 810-5, Rev. D, vol. I, sec. TCI-30 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 1, 1990.

$$T_{bwg,F3} = \left(1 - \frac{1}{L_{F3}}\right) T_p \tag{7}$$

$$T_p = 20 + 273.16 = 293.16 \text{ K} \tag{8}$$

$$T_{op,F1} = \frac{T_{sky}}{L_{F1} L_{wg}} + \frac{T_{bwg,F1}}{L_{wg}} + T_{tp} \tag{9}$$

$$T_{op,F2} = \frac{T_{sky}}{L_{F1} L_{F2} L_{wg}} + \frac{T_{bwg,F1}}{L_{F2} L_{wg}} + \frac{T_{bwg,F2}}{L_{wg}} + T_{tp} \tag{10}$$

$$T_{op,F3} = \frac{T_{sky}}{L_{F1} L_{F3} L_{wg}} + \frac{T_{bwg,F1}}{L_{F3} L_{wg}} + \frac{T_{bwg,F3}}{L_{wg}} + T_{tp} \tag{11}$$

Substitutions of Eqs. (2) through (8) into Eqs. (9) through (11) and algebraic manipulations lead to the loss factor expressions of

$$L_{F1} = \frac{1}{1 - \left[\dfrac{L_{wg}(\Delta T_{op})_{1g}}{T_p - T_{sky}}\right]} \tag{12}$$

$$L_{Fj} = \frac{1}{1 - \left[\dfrac{L_{F1} L_{wg} (\Delta T_{op})_{j1}}{T_p - T_{sky}}\right]} \tag{13}$$

where $j = 2$ or 3. The expressions for the differential operating noise temperatures are

$$(\Delta T_{op})_{1g} = T_{op,F1} - T_{op,g} \tag{14}$$

$$(\Delta T_{op})_{21} = T_{op,F2} - T_{op,F1} \tag{15}$$

$$(\Delta T_{op})_{31} = T_{op,F3} - T_{op,F1} \tag{16}$$

The above loss factor equations are functions of elevation angle, but measured T_{op} values are generally only available for the zenith test configuration. The following example is given for a zenith test configuration. Substitutions of Table 2 and Eq. (1) values into the above equations result in $L_{F1} = 1.02732$ (Cassegrain reflector associated dissipative losses), $L_{F2} = 1.0204$ (for BWG dissipative losses between F1 and F2), $L_{F3} = 1.04251$ (for BWG dissipative losses between F1 and F3, for F3 before mirror realignment), and $L_{F3} = 1.02686$ (for BWG dissipative losses between F1 and F3, for F3 after mirror realignment). As will be demonstrated in the following section, the above loss factors are useful for extracting the atmospheric noise temperatures from tipping-curve data obtained at the various BWG focal points.

C. Tipping Curves

Tipping-curve tests were performed at F1, F2, and F3 at various azimuth angles for the test package. A few of the tipping curves are shown in Fig. 7. Most of the tipping-curve data at 32 GHz were obtained at an azimuth angle of 50 deg, because, at this azimuth, the ground is at a very low (1-deg) elevation angle. For most large antennas in the DSN, it has not been possible to accurately extract atmospheric noise temperatures from tipping-curve data. However, it was reported in [2] that it was possible to extract X-band atmospheric values from tipping curves that agreed with theoretical values to ±0.2 K. A plausible explanation for this good agreement is the low scattering cross-sections of the newly designed slim tripod legs on this new BWG antenna.

For a flat-Earth model, the equation for extracting the zenith atmospheric noise temperature from tipping-curve data is derived as

$$T_{atmz} = L_{ant} \left[\frac{T_{op}(\psi_{el}) - T_{op}(90)}{\dfrac{1}{\sin(\psi_{el})} - 1}\right] + T'_{cb}\left(1 - \frac{1}{L_{atmz}}\right) \tag{17}$$

where

ψ_{el} = the elevation angle, deg

$T_{op}(\psi_{el})$ = the operating system temperature measured at ψ_{el}, K

$T_{op}(90)$ = the operating system temperature measured at $\psi_{el} = 90$ deg or at zenith, K

L_{atmz} = the loss factor for the zenith atmosphere. An estimate of this value can be made by using the loss factor for a standard atmosphere, or a better estimate can be obtained by using the SDSATM7M.BAS program[2]

T'_{cb} = the effective contribution to T_{op} from the cosmic background, K

L_{ant} = the loss factor from the Cassegrain antenna aperture to the input to the HEMT, K

[2] Courtesy of S. Slobin of the Jet Propulsion Laboratory, Pasadena, California. SDSATM7M.BAS is a modified version of SDSATM4.BAS, but gives the same answers.

The above equation should not be used for ψ_{el} values close to 90 deg or less than 10 deg. In practice, it has been found that Eq. (17) is most accurate when used with tipping-curve data in the region of a 30-deg elevation angle. For the special case of $\psi_{el} = 30$ deg, Eq. (17) becomes

$$T_{atmz} = L_{ant}[T_{op}(30) - T_{op}(90)] + T'_{cb}\left(1 - \frac{1}{L_{atmz}}\right) \quad (18)$$

At 32.0 GHz and under clear Goldstone Ka-band atmosphere conditions, the last term contributes only 0.05 K. For this BWG antenna, values of L_{ant} were calculated from values of L_{F1}, L_{F2}, L_{F3}, and L_{wg} given previously in Eqs. (12) and (13) and the following equations:

$$\text{For F1:} \qquad L_{ant} = L_{F1}L_{wg} \quad (19)$$

$$\text{For F2:} \qquad L_{ant} = L_{F1}L_{F2}L_{wg} \quad (20)$$

$$\text{For F3:} \qquad L_{ant} = L_{F1}L_{F3}L_{wg} \quad (21)$$

Substitution of values already presented in this article into Eqs. (19) through (21) gives: for F1, $L_{ant} = 1.09322$ (corresponding to 0.387 dB); for F2, $L_{ant} = 1.1155$ (corresponding to 0.475 dB); for F3 before mirror realignment, $L_{ant} = 1.1397$ (corresponding to 0.568 dB); and for F3 after mirror realignment, $L_{ant} = 1.1226$ (corresponding to 0.502 dB).

The reader is reminded that the L_{ant} values for the various focal points all contain the Cassegrain reflector loss factor $L_{F1} = 1.02732$ (corresponding to 0.117 dB) and $L_{wg} = 1.06414$ (corresponding to 0.27 dB). For the above values of L_{ant}, it was assumed that the antenna associated loss factors do not change significantly between zenith and 30-deg elevation angles, and therefore, the values for these loss factors could be derived from zenith conditions. For more accuracy, the dependence of these loss factors on elevation angle might have to be considered.

Table 4 is a worksheet showing the methodology used to extract atmospheric noise contributions from operating noise temperatures measured at zenith and 30-deg elevation angles at the various BWG focal points. Also shown is a comparison of the measured T_{atm} as obtained through the use of Eq. (18) with the computed T_{atm} as obtained through the SDSATM7M.BAS program, which is based on the weather data shown. Analysis of the results in Table 4 will show that the measured atmospheric values at 32 GHz are higher that the computed values by 2.4 + 0.5/–0.7 K.

For interest, the more rigorous method given by Eqs. (12), (13), (19), (20), and (21) was also used to extract X-band atmospheric noise temperatures from tipping curve data reported in [2]. It was found that the agreement between measured values and computed values at 8.45 GHz is now +0.1/–0.3 K, instead of the ±0.2 K deviations reported previously [2].

The tipping-curve results at X- and Ka-band indicate that either: (1) The theoretical weather model at Ka-band is not as good as at X-band for predicting atmospheric contributions from weather data, or (2) the tipping-curve data at Ka-band is corrupted more than at X-band by elevation-angle-dependent noise contributions from tripod leg scattering, subreflector spillover, main reflector spillover, and leakage through the antenna panels.

D. Subreflector Tests

Figure 8 shows a Ka-band subreflector defocus curve measured at F1 at an average elevation angle of 45 deg. To obtain this curve, the antenna tracked radio source 3C274, which had a source noise temperature of about 2.2 K at 32 GHz. Measurements of operating noise temperatures were made as functions of subreflector defocus positions. The Ka-band antenna-pointing model computer program was relied upon to keep the peak of the antenna beam pointed at the radio source. At the time the F1 subreflector curve measurements were made, a good experimental procedure had not yet been developed and, therefore, the experimental data are not very good for the sidelobe regions or for the main beam below −10 dB. However, it can be seen that over most of the main beam of the subreflector defocus curve, good agreement is obtained between a theoretical model and the experimental curve. The theoretical curve is based on a Geometric-Theory-of-Diffraction/Jacobi–Bessel antenna program.[3]

Figure 9 shows subreflector data measured at F3 for an average elevation angle of 45 deg. To obtain this curve, the antenna tracked radio source 3C84, which had a source noise temperature of about 4 K at 32 GHz at the time of these tests. To obtain data for this experimental curve, an improved test procedure was developed and implemented. Boresight measurements were performed more often (at least twice) during the subreflector test, which took about

[3] P. Cramer, "Calculated DSS-13 Subreflector Z-Axis Focus Curves, Feed at F1," Interoffice Memorandum 3328-90-0355 (internal document), Jet Propulsion Laboratory, Pasadena, California, September 20, 1990.

two hours to perform. Sufficient numbers of off-source and on-source measurements were taken to allow for corrections of measurement system drifts. Also shown in Fig. 9 is a theoretical defocus curve for 32 GHz and a 45-deg rigging angle. It can be seen in Fig. 10 that good agreement was obtained between Cramer predictions and the experimental data down to about −15 dB for the main beam. Good agreement was also obtained for some regions of the sidelobe patterns.

It should be pointed out that even though Cramer's predictions were made for F1 and the experimental results in Fig. 9 are for F3, there is reasonably good agreement between theory and experiment. For a properly designed and aligned BWG mirror system, the subreflector defocus gain curves at F3 should be nearly identical to those at F1. Therefore, a Z-defocus subreflector test at F3 provides a good method for verifying the integrity of the BWG system.

A Z-defocus subreflector test is also useful for revealing whether multipath signals exist within a large antenna system [7,8]. Scrutiny of the subreflector defocus data near the peak of the main beam (Fig. 10) indicates no unusual humps. The absence of humps and ripples on the main-beam defocus curve shown in Fig. 10 indicates that the magnitudes of any multipath signals within the BWG system are small (for a 100-MHz bandwidth measurement).

To obtain accurate subreflector defocus data below the 20-dB level (Fig. 9) while tracking a 4 K source, the radiometer needs to provide resolutions and accuracies of better than 0.04 K. In addition, the Ka-band antenna-pointing model should be good enough to allow blind pointing and remain on the peak of the source to within ±2 mdeg.

E. Overall Performance

The overall performance characteristics of the Ka-band test package are summarized in Table 5. The performance data are based on observed test data in a worst-case field environment, as well as on error analyses equations given in the Appendix to "Part I" [3] and also [9]. Examination of Table 5 reveals that the Ka-band test package performance was good in terms of resolution, gain stability, and linearity.

The test package was subjected not only to mechanical stresses during radio source tracking at various elevation angles at F1, but also to ambient temperatures ranging from 0 to 40 deg C and transportation to and from the ground locations F1, F2, and F3 several times. Test data in Table 1 showed that the long-term stability and repeatability of data for the various test configurations were very good.

IV. Concluding Remarks

The objective of determining the amount of degradations due to the BWG mirrors has been met. The Ka-band test package operated well in all of the different test configurations. Zenith operating noise temperatures, tipping-curve data, and subreflector defocus curves were successfully obtained at F1 and F3. As reported elsewhere, the Ka-band test package also enabled high-quality and reliable data to be obtained for antenna efficiency and antenna-pointing tests.

Acknowledgments

The assistance of J. Garnica, G. Bury, J. Crook, G. Farner, R. Littlefair, and R. Reese, all of DSS 13, is gratefully acknowledged. The authors also thank DSS 13 Group Supervisor, C. Goodson, and the DSS 13 Station Manager, A. Price, for their quick responses and superb management support for tasks involving special tests of the X- and Ka-band test packages.

References

[1] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable X-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part I: Design and Ground Tests," *TDA Progress Report 42-103*, vol. July–September 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 135–150, November 15, 1990.

[2] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable X-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part II: Tests on the Antenna," *TDA Progress Report 42-105*, vol. January–March 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 54–68, May 15, 1991.

[3] T. Y. Otoshi, S. R. Stewart, and M. M. Franco, "A Portable Ka-band Front-End Test Package for Beam-Waveguide Antenna Performance Evaluation—Part I: Design and Ground Tests," *TDA Progress Report 42-106*, vol. April–May 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 249–265, August 15, 1991.

[4] S. D. Slobin, T. Y. Otoshi, M. J. Britcliffe, L. S. Alvarez, S. R. Stewart, and M. M. Franco, "Efficiency Calibration of the DSS 13 34-Meter-Diameter Beam Waveguide Antenna at 8.45 and 32 GHz," *TDA Progress Report 42-106*, vol. April–May 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 283–297, August 15, 1991.

[5] L. S. Alvarez, "Initial Pointing Calibrations for the DSS 13 34-Meter Beam-Waveguide Antenna," *TDA Progress Report 42-106*, vol. April–May 1991, Jet Propulsion Laboratory, Pasadena, California, pp. 188–204, August 15, 1991.

[6] C. T. Stelzried and T. Y. Otoshi, "Radiometric Evaluation of Antenna-Feed Component Losses," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-18, no. 3, pp. 172–183, September 1969.

[7] G. S. Levy, D. A. Bathker, A. C. Ludwig, D. E. Neff, and B. L. Seidel, "Lunar Range Radiation Patterns of a 210-ft Antenna at S-band," *IEEE Transactions on Antennas and Propagation*, vol. AP-15, no. 2, pp. 311–313, March 1967.

[8] T. Y. Otoshi, "Multipath Tests on 64-m Antennas Using the Viking Orbiter-1 and -2 Spacecraft as Far-Field Illuminators," *DSN Progress Report 42-31*, Jet Propulsion Laboratory, Pasadena, California, pp. 41–49, February 15, 1976.

[9] C. T. Stelzried, "Operating Noise-Temperature Calibrations of Low-Noise Receiving Systems," *Microwave Journal*, vol. 14, no. 6, pp. 41–46 and 48, June 1971.

Table 1. Summary of Ka-band zenith operating noise temperatures at DSS 13 from October 12, 1990, through January 31, 1991

Config.	Observation Dates	Grand Average[a] T_{op}, K	Peak Deviations From Grand Avg., K
Ground	10/12/90, 11/09/90, 01/19/91, 01/31/91	84.7[b]	+1.6 −1.7
F1	10/13/90, 10/14/90, 01/11/91	91.8	+0.4 −0.6
F2	01/16/91, 01/17/91	97.0[c]	+0.4 −0.4
F3	11/10/90, 12/18/90	102.4	+0.1 −0.0
After mirrors and ellipsoid realigned on December 18, 1990			
F3	01/23/91, 01/25/91, 01/30/91	98.6[d]	+0.8 −0.7

[a] Normalized T_{op} values were computed through the use of Eq. (1).

[b] Ground values were reported in "Part I" [3]. The measured ground value of 84.7 K agrees closely with the predicted value of 84.5 K under standard conditions.

[c] No calibrations longer than 10 min were done at F2 with the antenna left at zenith.

[d] This number cannot be compared with the above F2 value. It is probable that the new F2 value is also lower after the mirror realignment, but a measurement was not made.

Table 2. Differential zenith operating noise temperatures for various test configurations at 32 GHz

Configurations Differenced	ΔT_{op}, K
F1 − Ground	7.1
F2 − F1	5.2
F3 − F1	10.6
After mirrors and ellipsoid realigned on December 18, 1990	
F3 − F1	6.8

Notes: See Table 1 for F1, F2, and F3 values.
Do not compare the values for F3 − F1 after December 18, 1990, with the value for F2 − F1, because the value at F2 might have become lower, but was not remeasured.

Table 3. Ka-band (32 GHz) measured zenith operating noise temperatures corrected for weather and waveguide noise changes

Config.	Observation Period	Average Measured T_{op}, K	Average Weather During Obsrv.	Computed T_{atm}, K	Computed L_{atm}, ratio	Physical WG Temp., deg C	T_{wg}, K	Normalized[a] T_{op}, K
F1	10/13/90 DOY 286 0046-0122 UT	92.1[b]	894.1 mb 22.8 °C 12.09% RH	6.83	1.0256 (0.1097 dB)	22.8	17.85	92.1
F1	10/14/90 DOY 287 0700-0800 UT	92.0	892.0 mb 19.9 °C 19.0% RH	7.62	1.0286 (0.1225 dB)	23.6	17.89	91.2
F1	01/11/91 DOY 011 0006-0900 UT	93.9	901.7 mb 5.2 °C 72.8% RH	9.56	1.0368 (0.1570 dB)	8.6	16.99	92.2
F2	01/16/91 DOY 016 0330 UT	97.45[c]	893.8 mb 11.6 °C 34.4% RH	7.98	1.0304 (0.1299 dB)	19.7	17.66	96.6
F2	01/17/91 DOY 017 1730-1832 UT	97.53[d]	902.9 mb 10.0 °C 28.5% RH	7.21	1.0274 (0.1176 dB)	18.4	17.58	97.4
F3	11/11/90 DOY 315 0300-0800 UT	102.2	901.5 mb 17.3 °C 15.5% RH	6.76	1.0255 (0.1095 dB)	20.7	17.72	102.4
F3	12/18/90 DOY 352 1200-1700 UT	102.0	896.1 mb 5.5 °C 31.6% RH	6.69	1.0256 (0.1099 dB)	17.4	17.52	102.5
			After mirrors and ellipsoid realigned on 91 DOY 018					
F3	01/23/91 DOY 023 0400-0500 UT	98.5	895.7 mb 6.3 °C 20.3% RH	5.97	1.0229 (0.0981 dB)	21.3	17.75	99.4
F3	01/25/91 DOY 025 0530-0600 UT	98.2	896.1 mb 8.0 °C 33.4% RH	7.22	1.0276 (0.1181 dB)	22.2	17.81	97.9
F3	01/30/91 DOY 030 0330-0600 UT	97.1	899.2 mb 4.1 °C 13.8% RH	5.40	1.0207 (0.0891 dB)	20.7	17.72	98.6

[a] Normalized T_{op} values were computed through the use of Eq. (1).

[b] Average zenith values during tipping curve tests.

[c] Extrapolated from 3C84 STAR track off source at 83.76 deg elevation.

[d] From rise-set tipping curve data at 120 and 50 deg azimuth.

Table 4. Measured and computed zenith atmospheric noise temperatures at Ka-band (32.0 GHz)

Config.	Azimuth Angle, deg	Observation Period	Average Weather	Computed L_{atmz}, ratio	L^a_{ant}, ratio	Meas. $T_{op}(30) - T_{op}(90)$, K	Meas.[b] T_{atm}, K	Computed[c] T_{atm}, K	ΔT_{atm}, K
F1	50.0	10/12/90 DOY 285 1746-1822 UT	894.1 mb 22.8 °C 12.1% RH	1.0256 (0.1097 dB)	1.0932 (0.387 dB)	8.11	8.92	6.83	2.09
F3	50.0	11/12/90 DOY 316 2213-2236 UT	901.4 mb 22.5 °C 13.8% RH	1.0270 (0.1158 dB)	1.1397 (0.568 dB)	8.46	9.69	7.22	2.47
F2	120.0	01/17/91 1991 DOY 17 1730-1800 UT	902.8 mb 9.8 °C 29.5% RH	1.0277 (0.1186 dB)	1.1155 (0.475 dB)	8.92	10.00	7.26	2.69
F2	50.0	01/17/91 1991 DOY 17 1803-1835 UT	902.9 mb 10.0 °C 28.0% RH	1.0273 (0.1169 dB)	1.1155 (0.475 dB)	9.01	10.10	7.16	2.94
After mirrors and ellipsoid realigned on 91 DOY 018									
F3	50.0	01/23/91 1991 DOY 23 1705-1748 UT	898.1 mb 9.3 °C 21.3% RH	1.0244 (0.1046 dB)	1.1226 (0.502 dB)	7.12	8.04	6.39	1.65

[a] See Eqs. (19) through (21).

[b] See Eq. (18).

[c] Computed values obtained from computer program SDSATM7M.BAS.

Table 5. Ka-band test package performance characteristics

Parameter	Performance Achieved
Receive polarization	Right circular polarization, left circular polarization (if reconfigured)
Receive frequencies	31.865–32.085 GHz determined from laboratory tests
T_{op} for test package on ground at DSS 13, zenith clear sky	85 K[a]
Measured T_{op} resolution for the above T_{op}, $\tau = 1$ sec, 100 MHz bandwidth	< 0.06 K
Gain stability over 0 to 40 deg C ambient temperature range	< 0.2 dB peak-to-peak, < 0.05 dB/hr (thermoelectric temperature control)
Total calibration nonlinearity error	< 2%
Radio source temp. T_s measurement accuracy (T_s is obtained from a Δ measurement)	$\pm[0.06 + 0.020 \times T_s]$ K for $2 < T_s < 10$ K
Overall T_{op} measurement accuracy,[b] K $10 < T_{op} < 150$ K	< 0.8 K peak-to-peak

[a] A major part of the total is due to 56.6 K from the HEMT and 17.7 K from waveguide losses.

[b] Based on error analysis, calibration errors, and estimated mismatch errors.

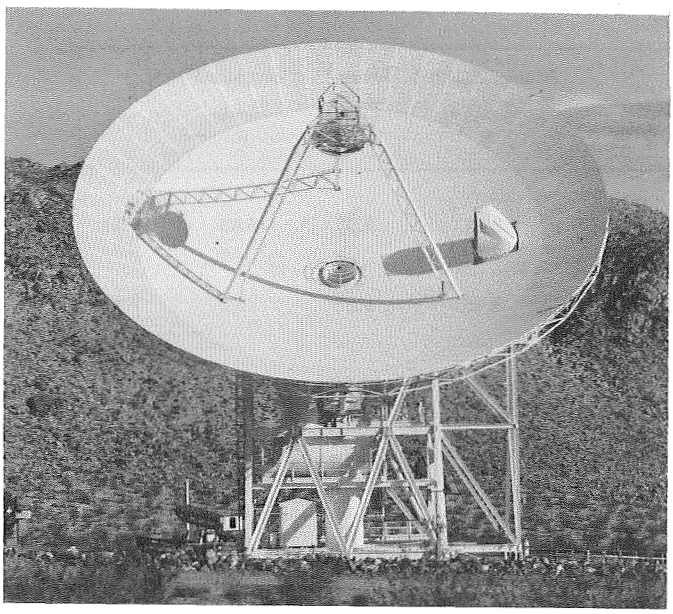


Fig. 1. The new constructed BWG antenna.

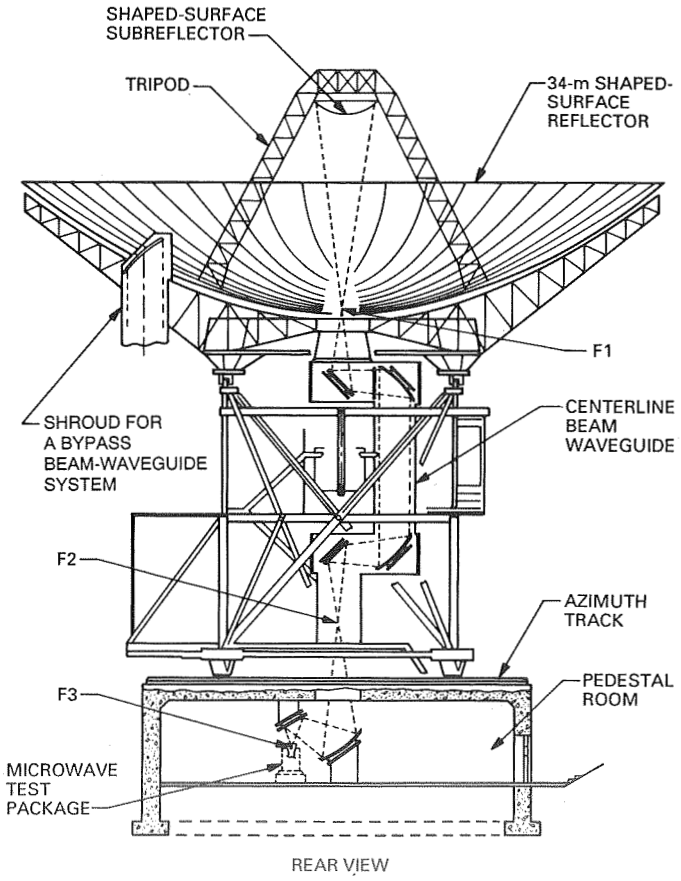


Fig. 2. The BWG antenna shown with focal points F1, F2, and F3.

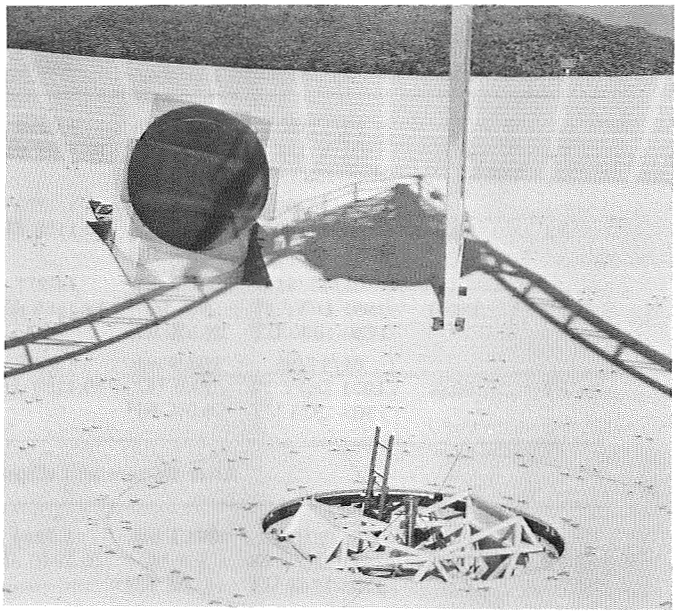


Fig. 3. Ka-band 29-dBi horn test package and mounting assembly installed at F1.

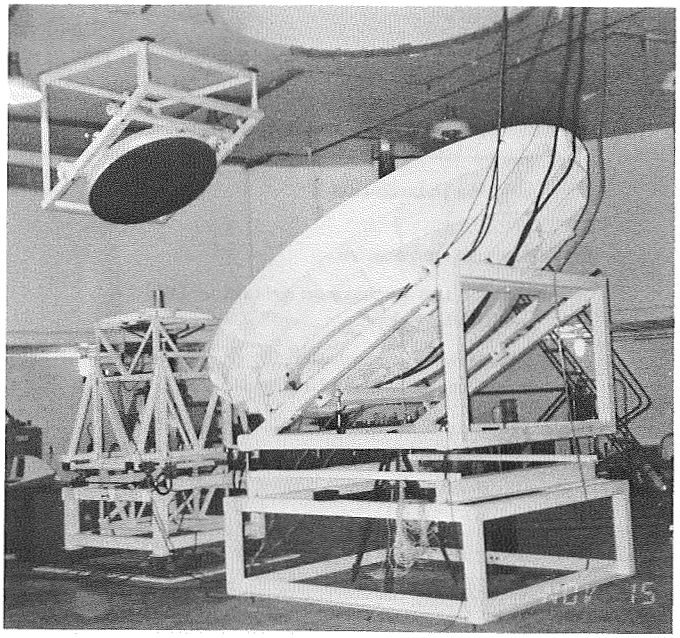


Fig. 4. Ka-band 23-dBi horn test package and mounting table installed at F3.

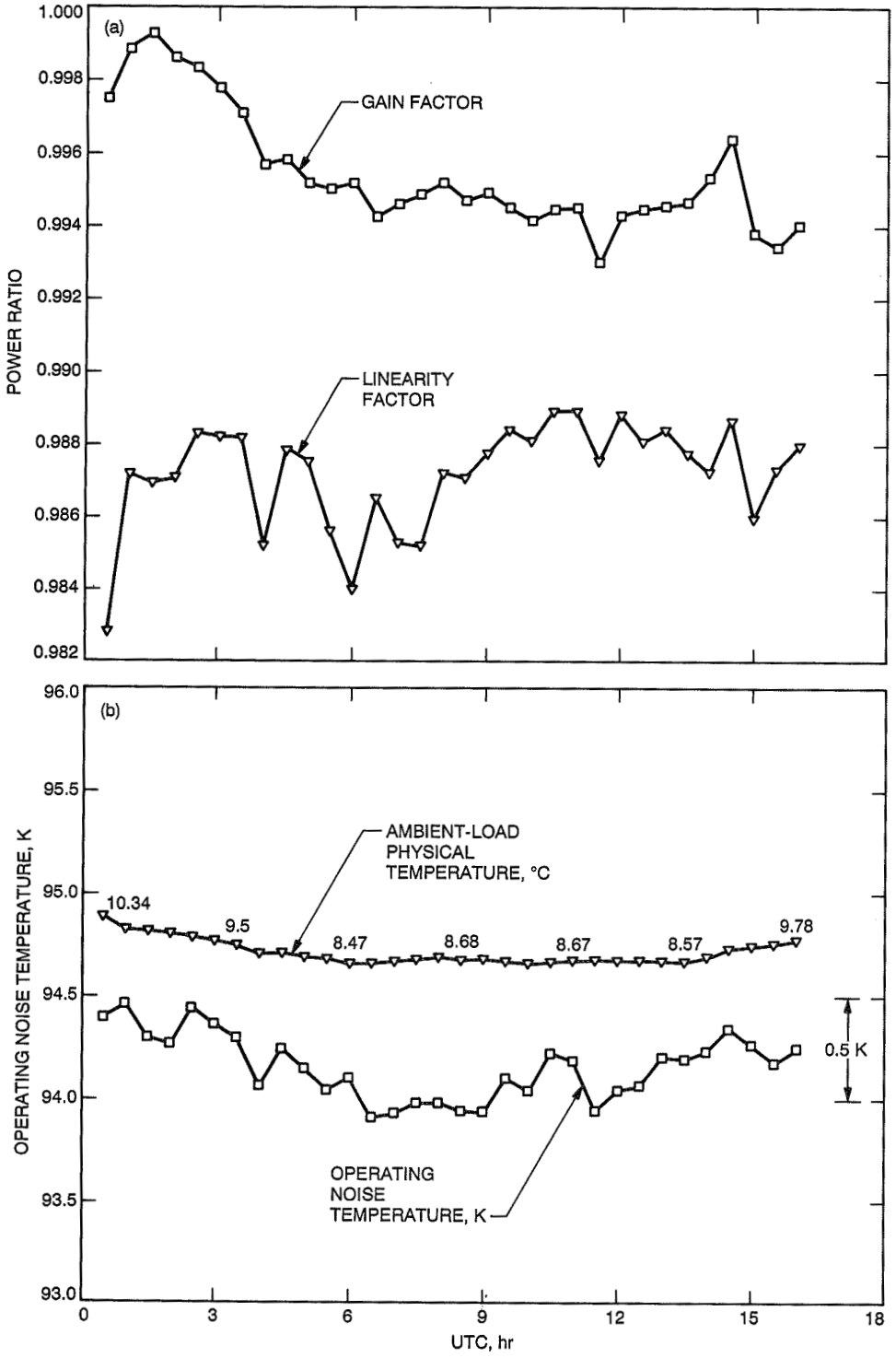


Fig. 5. Mini-cal data taken on January 11, 1991, with the Ka-band 29-dBi horn test package mounted at F1: (a) gain factor and linearity factor and (b) operating noise temperature and ambient load temperature.

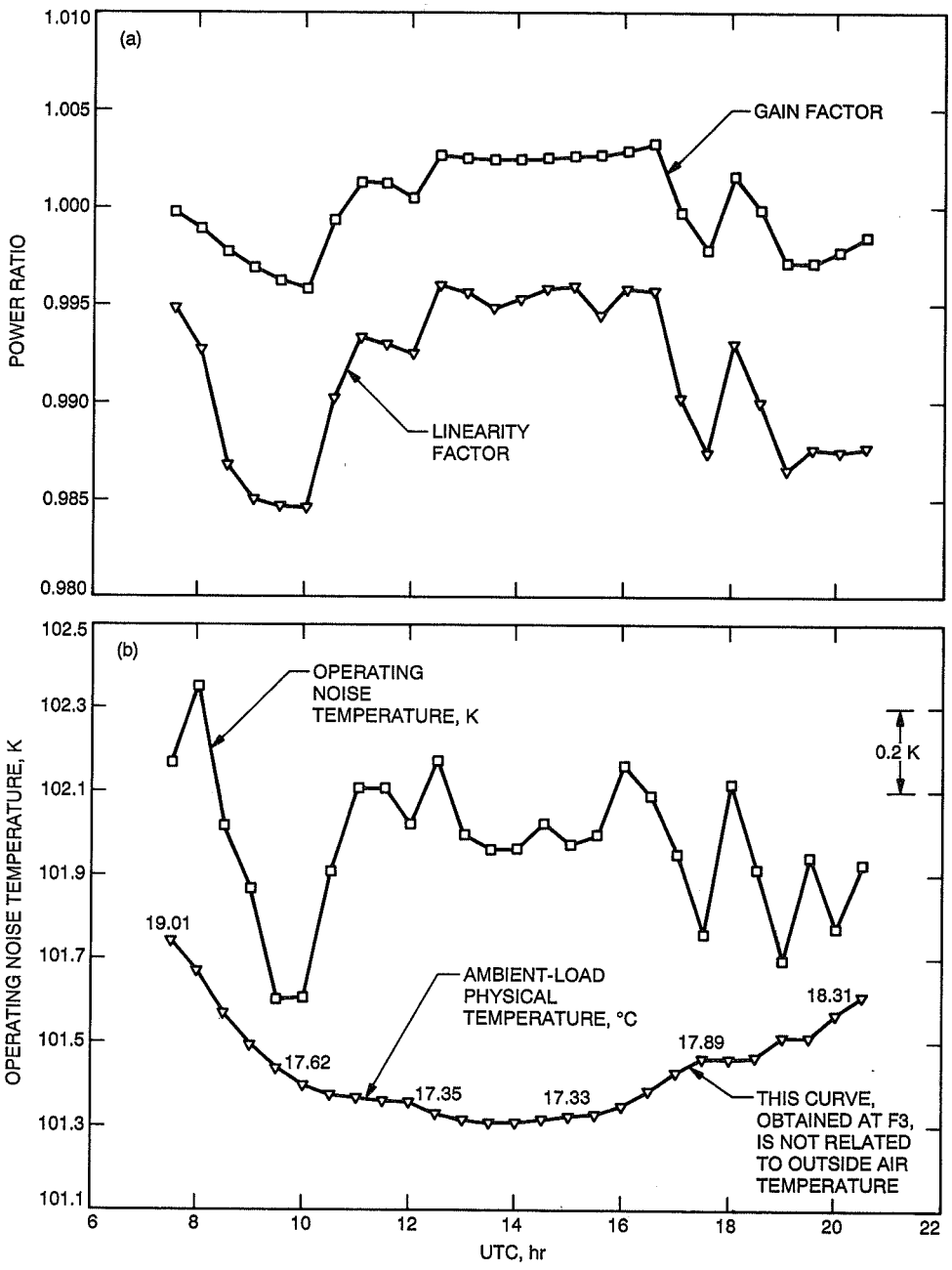


Fig. 6. Mini-calibration data taken on December 18, 1990, with the Ka-band 23-dBi horn test package mounted at F3: (a) gain factor and linearity factor and (b) operating noise temperature and ambient load temperature.

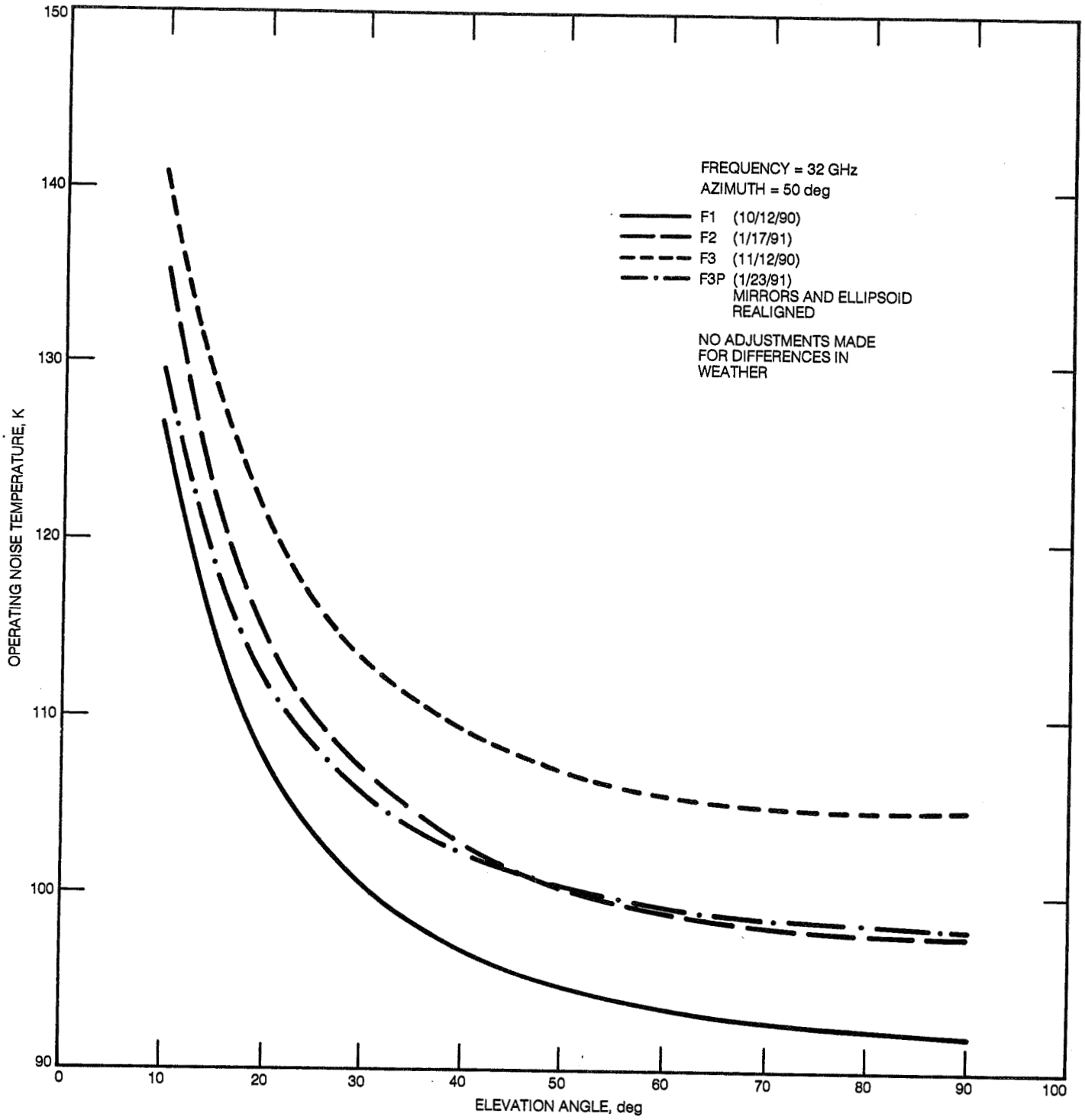


Fig. 7. Ka-band (32 GHz) tipping curves at F1, F2, and F3.

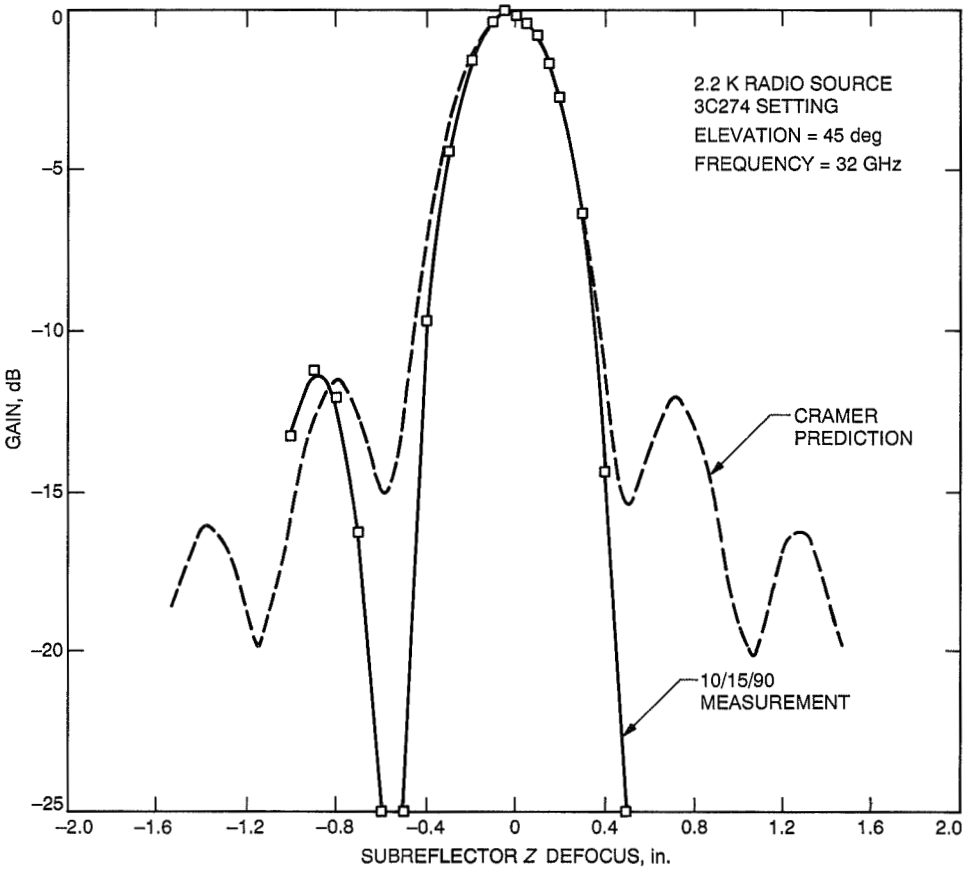


Fig. 8. Ka-band (32 GHz) subreflector *Z*-defocus curve measured at F1.

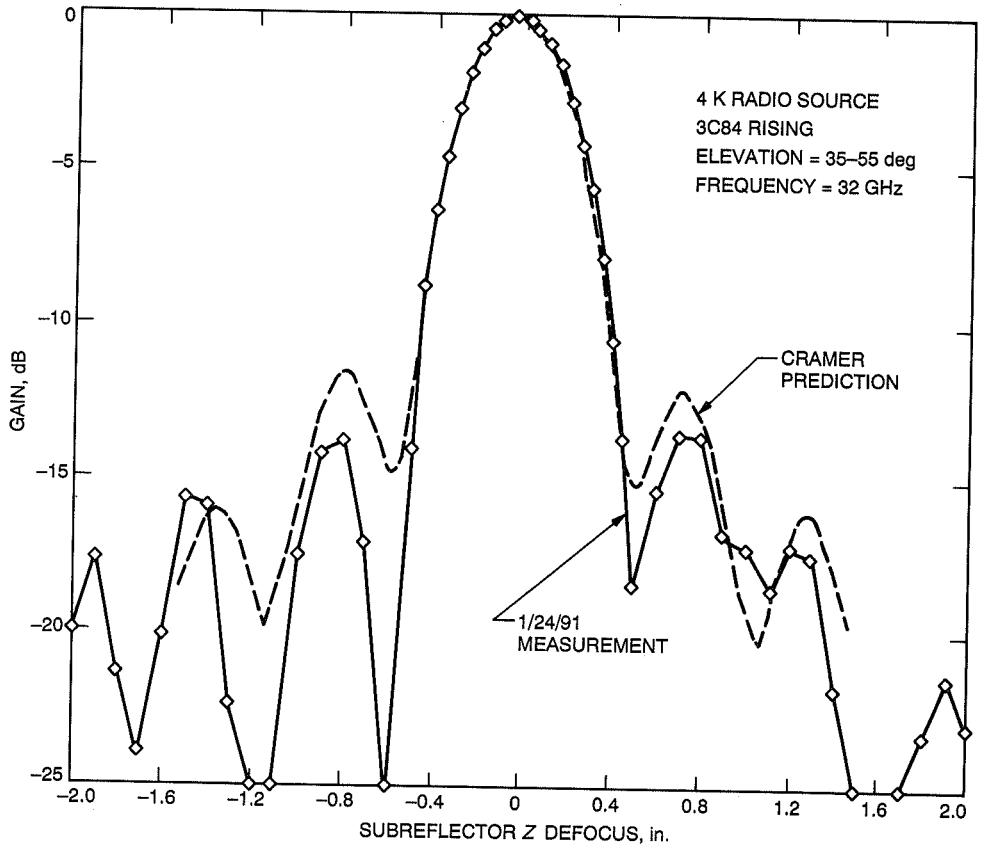


Fig. 9. Ka-band (32 GHz) subreflector *Z*-defocus curve measured at F3.

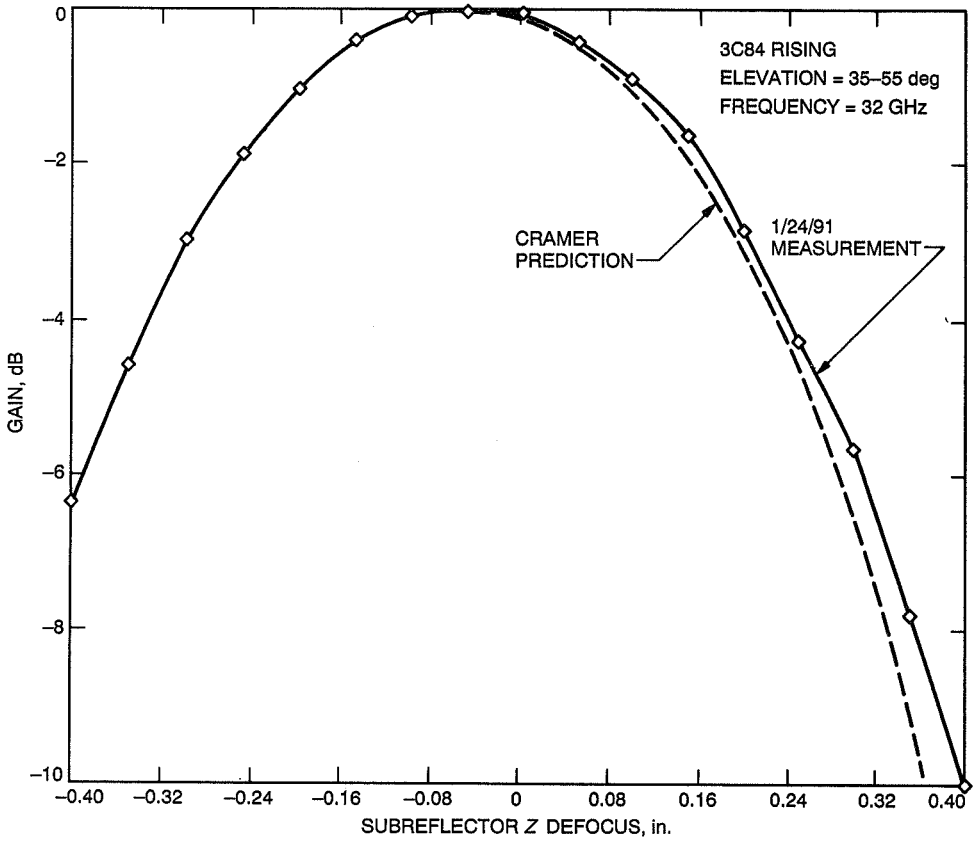


Fig. 10. Expanded detail of the main beam of the F3 Ka-band subreflector defocus curve (see Fig. 9).

Efficiency Calibration of the DSS 13 34-Meter Diameter Beam Waveguide Antenna at 8.45 and 32 GHz

S. D. Slobin
Telecommunications Systems Section

T. Y. Otoshi, M. J. Britcliffe, L. S. Alvarez, S. R. Stewart, and M. M. Franco
Ground Antennas and Facilities Engineering Section

Efficiency measurements at 8.45 and 32 GHz (X- and Ka-bands, respectively) have been carried out on the new 34-meter diameter beam waveguide antenna now in use at the NASA Goldstone Deep Space Communications Complex. The use of portable test packages enabled measurements at both the Cassegrain and beam waveguide focal points. Radio sources (quasars and Venus) were used as calibrators, and updated determinations of flux and source size correction were made during the period of the measurements. Gain and efficiency determinations as a function of elevation angle are presented, and the effects of the beam waveguide system and antenna structure are clearly seen. At the beam waveguide focus, an 8.45-GHz peak efficiency of 72.38 percent was measured; at 32 GHz, 44.89 percent was measured.

I. Introduction

From July 1990 through January 1991, the new 34-meter diameter beam waveguide (BWG) antenna at the NASA Goldstone Deep Space Communications Complex in the Mojave Desert was tested as part of its post-construction performance evaluation. This research and development antenna, designated Deep Space Station 13 (DSS 13), is the prototype for a new generation of Deep Space Network (DSN) antennas utilizing the BWG-type of antenna feed system. Figure 1 shows the mechanical and microwave optics designs of the antenna.

Efficiency and pointing performance were characterized at 8.45 and 32 GHz (X- and Ka-bands, respectively) at both the Cassegrain (f1) and BWG (f3) focal points. The f1 focal point is located close to the vertex of the main reflector, while the f3 focal point is located about 35 meters and six additional reflectors away in a subterranean pedestal room. The BWG type of design enables the simultaneous use of numerous different feeds in the relatively well-controlled conditions of the pedestal room, as contrasted with the present DSN configuration with feedhorns, waveguide components, and low-noise amplifiers installed above the surface of the dish, where maintenance and modification tasks are considerably more difficult. In addition, increased system noise temperature due to rain on the feedhorn covers and dichroic plates has been eliminated.

The performance evaluation methods described in this article are unique in that a direct experimental measurement was made of the antenna efficiency and gain degradation caused by the BWG mirror system. The tests in-

volved the use of X- and Ka-band portable test packages that were installed at either focal point f1 or f3. To the authors' knowledge, this is the first known use of portable test packages to determine gain and efficiency at various locations in the microwave optics path of an antenna.

Section II of this article presents the boresight measurement technique and pointing correction methodology used in looking at calibration radio sources of both known and unknown flux and source size correction values. Section III describes the methodology of antenna efficiency determination. Section IV presents methods for atmospheric attenuation correction. Section V presents characteristics of the radio sources used during the boresight measurements. The last sections present the results of the measurements and final deduced values of flux and source size correction for the not-well-known calibration radio sources.

II. Boresight Technique

Critical to the proper performance of the measurements was the use of a boresighting procedure to accurately point the antenna at the radio sources used in the calibration process. The method used was a seven-point boresight technique. This method moved the antenna sequentially in both the cross-elevation (XEL) and elevation (EL) directions both on and off the source. In each direction, the antenna was positioned off-source ten half-power (one-sided) beamwidths; one half-power beamwidth at the 3-dB point; approximately 0.576 half-power beamwidth at the 1-dB point; on source; and then similar offsets on the other side. For example, at X-band with the 34-meter antenna, the full 3-dB beamwidth is 65 mdeg. The offsets used were 325, 32.5, 18.7, and 0 mdeg in each direction. At Ka-band, with a full 3-dB beamwidth of 17 mdeg, the offsets were 85, 8.5, 4.9, and 0 mdeg.

For each scan, an off-source baseline is generated from the two off-source points. A Gaussian curve (relative to the baseline) is fitted to the five remaining on-source points and the peak value of the curve is calculated. Also, the position of the peak is calculated as a measure of the pointing error for that scan. One pair of scans (one XEL and one EL) is considered to be one measurement or data point. The seven pointing offsets for each new scan are corrected for pointing errors found from the previous similar scan, in order to maintain pointing throughout a track. As an example, if for a particular a priori pointing model installed in the antenna control subsystem it is found that from boresight-to-boresight a consistent +3-mdeg pointing error is found in the elevation scans, then this correction is made from scan-to-scan in order to maintain pointing. After 10 boresights, a 30-mdeg total pointing error relative to the model will have been found, although the scan-to-scan pointing error will not exceed 3 mdeg. For the purposes of the efficiency measurements, it is thus assumed that perfect pointing is maintained for all scans and that the small *calculated* pointing errors are due to random errors in the system operating noise temperature measurements, rather than due to *actual* mispointing. No efficiency corrections for any small pointing errors were made in these measurements. Obviously bad data points were discarded.

Two data files are generated during the measurements: an efficiency file and a pointing file. The efficiency file consists of time, antenna azimuth and elevation, and XEL and EL estimates of half-power beamwidth, peak-source noise temperature contribution, and scan pointing error. The pointing file consists of time, antenna azimuth and elevation, and pointing correction. Pointing correction is a cumulative value calculated from the sum of the pointing errors determined from the previous boresights. For example, if the first three boresights found +3, +2, and +4 mdeg pointing errors (relative to the a priori pointing model) in the elevation scans, the total accumulated pointing correction needed after the third boresight would be −9 mdeg. This correction is used in the generation of a new systematic pointing-error correction model for use in future measurements. With a perfect pointing-error correction model, the scan-to-scan pointing errors would be small and of random sign.

This boresight technique was developed at the Jet Propulsion Laboratory as a computer-assisted interactive program utilizing simultaneous operation of a personal computer and the antenna local control and display (LCD) console for pointing the antenna, determining and entering pointing offsets, entering noise temperature data, calculating radio source peak-noise contribution, and determining antenna pointing errors.

III. Antenna Efficiency Determination

For a radio source of known flux density, the increase in system noise temperature as determined by the boresight measurements is a measure of the antenna efficiency. As defined here, the efficiency is referenced to the input of the low-noise amplifier (cooled high-electron-mobility transistor, HEMT) and includes the losses of the feed system. Alternative methods specify efficiency at the aperture of the feedhorn or at the antenna aperture itself. The term aperture efficiency refers to antenna gain relative to that of a uniformly illuminated circular aperture having the same diameter as the antenna, e.g., a 70-percent efficient antenna has 1.549-dB less peak gain than does the circular aperture.

The radio-source noise temperature increase measured by the antenna is given by [1]:

$$\Delta T = \frac{\eta S A}{2kC_rC_p}$$

where

η = antenna aperture efficiency

S = radio source flux, watts/m^2/Hz

A = antenna area, m^2

k = Boltzmann constant, 1.38062 × 10^{-23} J/K (Wsec/K)

C_r = source size correction, typically 1.0 for point sources, up to ~1.5 for extended sources, including planets

C_p = pointing correction, assumed = 1.0

The flux, S, is typically given in units of "Janskys," where 1 Jansky (Jy) = 1 × 10^{-26} W/m^2/Hz.

Measured ΔT is compared with the quantity $[(SA)/(2kC_r)]$ to give antenna efficiency. Thus:

$$\eta = \frac{\Delta T}{SA/2kC_r}$$

$$= \frac{\Delta T}{T100/C_r}$$

$T100$ is what would be measured by a perfect antenna looking at a point source emitting the same flux as the observed radio source. C_r is a function of the source structure at a particular frequency and the antenna pattern of a particular antenna at that frequency. For a particular antenna, frequency, and radio source, the $T100/C_r$ can thus be specified.

For a planet, the flux is determined from the blackbody disk temperature and the angular size of the source (which changes as the distance between Earth and the planet changes).

The flux is given by [2]:

$$S = \frac{2kT\Omega}{\lambda^2}$$

where

k = Boltzmann constant

T = blackbody disk temperature, K

λ = wavelength, m

Ω = solid angle subtended by source = $(\pi D^2)/(4R^2)$

D = planet diameter, km

R = distance to planet, km

IV. Correction for Atmospheric Attenuation

ΔT is an on–off source measurement, and the Earth's atmosphere attenuates the true source contribution that would be measured under vacuum conditions. The total atmospheric attenuation was estimated from surface weather conditions during all measurements. The surface temperature, pressure, and relative humidity at the site were recorded every half hour. Typical zenith values of attenuation at Goldstone under average clear-sky conditions are:

$$\text{X-band} : A_{zen} = 0.035 \text{ dB}$$

$$\text{Ka-band} : A_{zen} = 0.115 \text{ dB}$$

The attenuation in decibels at elevation angle θ is modeled as:

$$A(\theta) = \frac{A_{zen}}{\sin(\theta)}$$

The loss factor at that elevation angle is:

$$L(\theta) = 10^{A(\theta)/10}$$

The "vacuum ΔT" then becomes:

$$\Delta T = L(\theta)\Delta T \text{ measured}$$

V. Radio Source Flux Values and Size Corrections

At both X- and Ka-bands, radio source 3C274 (Virgo A) is considered a principal calibrator, although at Ka-band, Venus is considered well-known enough to be used as a primary source also. At both X- and Ka-bands, the principal calibrators are used to establish the peak antenna efficiency near the antenna rigging angle (40–50 deg

elevation). As these sources typically do not rise to high elevation angles, other sources which do are also used to establish the shape of the efficiency curve with the elevation angle. High declination sources, such as 3C84 and 3C123, are used for this purpose. As their fluxes are not well-known (or the sources are variable), the efficiency curves generated are normalized to the peak values determined by the principal calibrators.

Table 1 summarizes the flux, size correction, and other factors associated with the radio source calibrators used during the DSS 13 X-band calibration measurements. It should be noted that radio sources 3C84 and 3C273 are known to be variable; however they are point sources and are useful for pointing model development.

VI. Efficiency Measurements: X-Band at f1, September 1990

By using the peak efficiency value determined by 3C274 to define the antenna performance, it was found that the efficiencies as determined by the other sources (using the $T100/C_r$ values above) varied by as much as 40 percent. Adjustments were made separately to the data points of each source to normalize their curve-fitted peak values to the 3C274 curve-fitted peak value. Table 2 gives the required efficiency adjustments and the deduced flux and $T100/C_r$ values thus obtained. In the case of point sources (3C84, 3C273), the adjustment reflects a flux error. For non-point or extended sources (DR21, 3C123) the adjustment could be due to either flux or source size-correction errors (which are at present unable to be separated); thus only the deduced $T100/C_r$ value is given. It should be noted that the variable sources (3C84, 3C273) may vary considerably over a period of weeks, thus the $T100/C_r$ values found represent, at best, interim values.

The zenith atmosphere attenuation corrections made during these measurements ranged from 0.036 dB to 0.046 dB, during clear weather. The X-band peak efficiency of the DSS 13 antenna at f1 was determined to be 75.35 percent at an elevation angle of 45.63 degrees. This corresponds to a peak gain of 68.34 dBi at 8450 MHz.

The above efficiency value was measured in the post-holographic condition of the antenna, with the main reflector panels in their final adjustment condition. After initial acceptance of the antenna in July 1990, preliminary X-band f1 efficiency measurements were made. These measurements yielded a peak efficiency of 71.88 percent. Thus, the panel adjustment increased efficiency by 3.5 percentage points (a 0.21-dB improvement) in addition to resetting the rigging angle to near 45 deg instead of the initial 58 deg. (These values are shown in Table 8, along with the results of all other post-holographic measurements.) Figure 2 shows the pre- and post-holographic comparison. The tolerances shown in Fig. 2 include the effects of random measurement error and curve fitting methodology only. Other effects which are common to both sets of measurements (source flux and size correction) are not included, so as to emphasize the fact that the 0.21-dB improvement shown is indeed statistically significant. Figure 3 shows the X-band f1 efficiency curve along with the adjusted data points.

VII. Efficiency Measurements: X-Band at f3, November 1990

By methods similar to those described above, X-band f3 measurements were made two months later. The radio sources used were restricted to three: 3C274, 3C84, and 3C123. Atmospheric attenuation at zenith varied from 0.035 dB to 0.037 dB, a lower and smaller range than in September, undoubtedly due to lower water vapor contribution. By using 3C274 as the standard gain calibrator and the other two sources to provide shape information, it was found that the efficiency of the beam waveguide system was somewhat lower than at f1. The final peak efficiency was determined to be 72.38 percent at an elevation angle of 40.21 deg, a decrease of 2.97 percent. The corresponding f3 gain is 68.17 dBi. Table 3 gives the deduced flux and $T100/C_r$ values determined during the f3 measurements.

It is seen that the $T100/C_r$ value of 3C123 appears to have decreased 1.51 percent, and that of 3C84 appears to have decreased 2.12 percent in the two months between September and November 1990. Again, these changes are relative to 3C274, and because they are both small, they may not be significant. As 3C123 is not considered to be variable, the 1.51 percent change may be due to noise in the data.

Figure 3 shows the adjusted data points and individual curve fits for the three sources (3C123 and 3C84 adjusted). It appears that the fits for 3C123 and 3C84 are significantly different at high elevation angles (at meridian transit). Because of the different declinations of these two sources, 3C123 passes south at meridian transit and 3C84 passes north, although both are only 6 deg from zenith. It is now believed that the different efficiency values at high elevation are azimuth-related and are due to a misalignment of the ellipsoidal and flat mirrors at the bottom of the beam waveguide system. At meridian transit, these two sources have 180 deg different azimuths. Note that

3C274, which also passes south of the antenna at meridian transit (although at only 67 deg elevation), has a curve fit more like that of 3C123, which also passes south.

VIII. Efficiency Measurements: Ka-Band at f1, October 1990

The four radio sources used for Ka-band measurements were 3C274, 3C84, 3C273, and Venus. The 3C274 flux appears to be well known, as determined in a recent analysis at JPL.[1] The source size correction was calculated by integrating the antenna beam (0.017-deg beamwidth) over a frequency-extrapolated map of the source. The source-size correction thus determined was 1.273. The flux of 3C84 was unknown, but a starting value was used based on 32-GHz measurements carried out on the Goldstone 70-meter antenna (DSS 14) in 1989 [5]. In that report, a flux of 43.71 Janskys was deduced with reference to measurements made by using 3C274 and Venus as standard calibrators. It was found during the f1 measurements that this value required substantial adjustment to bring the October 1990 efficiency measurements into agreement with those determined by Venus and 3C274.

Flux from Venus was calculated from the expression given above in Section III. The source size correction, C_r, for a disk is calculated from:[2]

$$C_r = \frac{X}{1 - e^{-X}}$$

where

$X = (r^2)/(2\sigma^2)$

r = angular radius of disk, deg

$\sigma = \Theta_o/2.3548$, deg

Θ_o = full 3-dB antenna beamwidth, deg

$T100/C_r$ is calculated as usual from

$$\frac{T100}{C_r} = \frac{SA}{2kC_r}$$

as described above.

A value of 475 K was used for the blackbody disk temperature of Venus [5,6]. The planetary disk diameter of Venus used was 12,240 km [6]. During the Ka-band f1 measurement period, Venus was near superior conjunction, and the distance from Earth to Venus changed very little, from about 1.70 to 1.72 AU. Table 4 gives the values used in the Ka-band calibration at f1.

Using the Table 4 $T100/C_r$ values for the sources listed, the f1 efficiency as determined by using Venus, was found to be 51.50 percent. By using 3C274, the efficiency was found to be 53.12 percent. As both sources are regarded to be standard calibrators, the results were averaged, yielding a final f1 efficiency value of 52.31 percent at an elevation angle of 44.82 deg. The corresponding gain is 78.33 dBi. The zenith atmospheric attenuation during the measurement period ranged from 0.133 dB to 0.233 dB. Adjustments in the flux and $T100/C_r$ values of 3C84 and 3C273 were made in order to match the efficiency found above. Table 5 gives these adjustments.

It should be noted that the adjustments for 3C274 and Venus are shown as needed only to establish an average peak gain value. As the sources are considered to be standard calibrators, the adjustments do not imply new flux or $T100/C_r$ values.

Figure 4 shows the adjusted data points and the curve fit for the Ka-band f1 measurements.

IX. Efficiency Measurements: Ka-Band at f3, January 1991

Ka-band measurements at f3 during December 1990 showed a very apparent efficiency anomaly, which unfortunately became known as the "hysteresis effect," because it was initially thought to be due to a mechanical inelasticity (which turned out not to be the case). After numerous tests, it was determined that the most likely cause of this effect was a misalignment of the ellipsoidal mirror in the pedestal room. Realignment of the mirror was done in early January 1991. It was then decided to re-do the Ka-band f3 measurements throughout that month. Only Venus, 3C274, and 3C84 were used for these tests. Table 6 gives the flux and $T100/C_r$ values used during the Ka-band f3 measurements. Note that a dummy value of $T100/C_r = 10.000$ is used for 3C84 (based on 9.629 deduced at f1). The real value would be determined after comparison with the Venus and 3C274 efficiency results.

Zenith atmospheric attenuation during these tests varied from 0.096 dB to 0.120 dB, substantially lower than during the f1 tests. Weather during the tests was clear and very cold, with very low absolute humidity. Table 7

[1] P. H. Richter, *Virgo-A Flux Density*, JPL Interoffice Memorandum 3393-90-137 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 18, 1990.

[2] P. H. Richter, personal communication, Jet Propulsion Laboratory, Pasadena, California, October 1, 1990.

gives the efficiency adjustments needed and the deduced values of flux and $T100/C_r$. Note that although adjustments are shown for Venus and 3C274, the given values of flux and $T100/C_r$ are not changed from the initial values given above, as they are considered to be calibration standards.

The peak Ka-band f3 efficiency was determined to be 44.89 percent at an elevation angle of 45.41 deg. This corresponds to a gain of 77.66 dBi.

Several things should be noted about Table 7. Compared with the Ka-band f1 measurements, the 3C274 and Venus adjustment directions have been reversed. The adjustment for 3C84 is based on an assumed $T100/C_r$ value of 10.000, a dummy value. The final deduced flux and $T100/C_r$ values for 3C84 vary only by 1.24 percent from those determined in the f1 measurements, even though the source is variable and three months time had elapsed. This may hint at the period of variation of the source.

Figure 4 shows the adjusted data points and the curve fits for the three calibration sources. Again, the spread of curve fits at the high elevation angles shows the effect of possible ellipsoidal mirror misalignment. Note that the mirror alignment for the Ka-band f3 tests is different from that for the X-band f3 tests in November 1990. As compared with the X-band curve fits, the southern-passing sources (3C274, Venus) straddle the curve fit of the northern-passing source (3C84), rather than being above it. Because of the low declination of Venus during the f3 tests (resulting in low peak-elevation angle), the curve fit at high elevation angles may be in error, thus giving this result.

X. Final Results

Table 8 gives a summary of the final gain and efficiency measurements described above. In addition, the estimated 1-σ errors are given, as well as shown in Figs. 3 and 4. The errors are due to source flux uncertainty, source size-correction error, data noise, curve fitting, and reduction methodology. For purposes of comparison, Fig. 5 shows all data sets and curve fits together. The frequency-dependent effect of main-reflector structural deformation is clearly seen in this figure.

Table 9 gives the coefficients of the final antenna efficiency curve-fit polynomials used in the generation of Figs. 3, 4, and 5.

Table 10 gives final radio-source flux and $T100/C_r$ values deduced from the efficiency measurements described above. When two different values have been found from f1 and f3 measurements, the value presented here is the average of the two. These values are, at least, a starting point for future measurements. In the case of standard calibrators (3C274, Venus), no new values are given. The $T100/C_r$ values given for non-point sources are valid only for use in the calibration of 34-meter diameter antennas with X- and Ka-band beamwidths of 65 and 17 mdeg, respectively. The C_r values were calculated by using these beamwidths. Care should be taken in calibration measurements when the antenna beamwidth varies from these numbers. Simple modeling is typically not accurate.

XI. Conclusion

This article has presented a complete review of the initial 8.45- and 32-GHz efficiency calibration of the new NASA Deep Space Network beam waveguide antenna at the Goldstone Deep Space Communications Complex. The novel technique of using portable test packages to evaluate antenna performance at various locations in the microwave optics path has proved to be straightforward to implement and successful in operation. At the beam waveguide focus, 8.45- and 32-GHz peak efficiencies of 72.38 and 44.89 percent, respectively, were achieved. The efficiency determinations presented here met the functional requirements of the DSS 13 project and agreed well with predictions which considered feed illumination and spillover, waveguide and mirror losses, and beam waveguide effects.

Acknowledgments

The authors wish to thank R. L. Riggs for the development of the boresight methodology and programming, without which the measurements described here would have been impossible. D. L. Brunn assisted greatly with experiment planning. P. H. Richter provided updated radio-source flux and source size-correction values. The authors are also indebted to Goldstone DSS 13 personnel for their support during the many months of the measurement program.

References

[1] J. A. Turegano and M. J. Klein, "Calibration Radio Sources for Radio Astronomy: Precision Flux Density Measurements at 8420 MHz," *Astron. Astrophys.*, vol. 86, pp. 46–49, 1980.

[2] J. A. Turegano and M. J. Klein, "Precision Flux Density Measurements of the Giant Planets at 8420 MHz," *Astron. Astrophys.*, vol. 94, pp. 91–94, 1981.

[3] P. H. Richter and S. D. Slobin, "DSN 70-Meter Antenna X- and S-Band Calibration, Part I: Gain Measurements," *TDA Progress Report 42-97*, vol. January–March 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 314–351, June 15, 1989.

[4] P. H. Richter and S. D. Slobin, "Errata to DSN 70-Meter Antenna X- and S-Band Calibration, Part I: Gain Measurements," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, p. 220, November 15, 1989.

[5] M. S. Gatti, M. J. Klein, and T. B. H. Kuiper, "32-GHz Performance of the DSS 14 70-Meter Antenna: 1989 Configuration," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 206–219, November 15, 1989.

[6] P. G. Steffes, M. J. Klein, and J. M. Jenkins, "Observations of the Microwave Emission of Venus from 1.3 to 3.6 cm," *Icarus*, vol. 84, pp. 83–92, 1990.

Table 1. Radio sources used for 8.45-GHz (X-band) calibrations at f1 and f3, September 1990 and November 1990

Source	Declination, J2000.0	Peak Elevation at DSS 13 LAT = 35.25 deg	Flux, Jy	C_r	$T100/C_r, K$
3C274[a]	12.391	67.1	44.555	1.087	13.477
3C123[b]	29.671	84.4	9.404	1.0054	3.0753
DR21[b]	42.329	82.9	19.902	1.0085	6.489
3C84[c,d]	41.512	83.7	45.79	1.000	15.056
3C273[c]	2.052	56.8	36.46	1.000	11.988

[a] P. H. Richter, *Virgo-A Flux Density*, JPL Interoffice Memorandum 3393-90-137 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 18, 1990.

[b] P. H. Richter, personal communication, Jet Propulsion Laboratory, Pasadena, California, July 26, 1990.

[c] The values as shown in [d] reduced by 0.076 dB as per [3] calibration analysis of 70-meter antenna gain, and [4].

[d] M. Klein, et al., *DSN Radio Source List for Antenna Calibration*, JPL Document D-3801, Rev. B (internal document), Jet Propulsion Laboratory, Pasadena, California, September 25, 1987.

Table 2. Efficiency adjustments, deduced flux, and *T100/Cr* values: 8.45 GHz (X-band) at f1, September 1990

Source	Efficiency Adjustment Needed	Deduced Flux, Jy	Deduced $T100/C_r, K$
3C274[a]	1.00000	44.555	13.477
3C123[b]	0.98722	—	3.115
DR21[b]	0.99478	—	6.523
3C84[c]	1.40141	32.67	10.743
3C273[c]	1.24420	29.30	9.635

[a] Extended

[b] Nonpoint

[c] Variable, point

Table 3. Efficiency adjustments, deduced flux and *T100/Cr* values: 8.45 GHz (X-band) at f3, November 1990

Source	Efficiency Adjustment Needed	Deduced Flux, Jy	Deduced $T100/C_r, K$
3C274[a]	1.00000	44.555	13.477
3C123[b]	1.00240	—	3.068
3C84[c]	1.43183	31.98	10.515

[a] Extended

[b] Nonpoint

[c] Variable, point

Table 4. Radio sources used for 32 GHz (Ka-band) calibrations at f1, October 1990

Source	Declination, J2000.0	Distance, AU	Peak Elevation at DSS 13 LAT = 35.25 deg	Flux, Jy	C_r	$T100/C_r, K$
3C274[a]	12.391		67.1	16.22	1.273	4.190
3C84[b]	41.512		83.7	43.71	1.000	14.372
3C273[b]	2.052		56.8	29.54	1.000	9.713
Venus 1990[c]						
Oct 01	+1.6	1.69539	56.4	27.323	1.00828	8.914
Oct 10	−2.9	1.70591	51.9	26.987	1.00818	8.806
Oct 20	−7.9	1.71302	46.9	26.764	1.00811	8.733
Oct 30	−12.5	1.71538	42.3	26.690	1.00809	8.709

[a] P. H. Richter, *Virgo-A Flux Density*, JPL IOM 3393-90-137 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 18, 1990.

[b] M. S. Gatti, M. J. Klein, and T. B. H. Kuiper, "32-GHz Performance of the DSS 14 70-Meter Antenna; 1989 Configuration," *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 206–219, November 15, 1989.

[c] Declination of date

Table 5. Efficiency adjustments, deduced flux, and *T100/Cr* values: 32 GHz (Ka-band) at f1, October 1990

Source	Efficiency Adjustment Needed	Deduced Flux, Jy	Deduced $T100/C_r, K$
3C274[a]	0.98475	16.22	4.190
Venus[b]	1.01573	Table 4	Table 4
3C84[c]	1.49278	29.281	9.628
3C273[c]	0.94022	31.418	10.331

[a] Extended

[b] Nonpoint

[c] Variable, point

Table 6. Radio sources used for 32-GHz (Ka-band) calibrations at f3, January 1991

Source	Declination, J2000.0	Distance, AU	Peak Elevation at DSS 13 LAT = 35.25 deg	Flux, Jy	C_r	$T100/C_r, K$
3C274[a]	12.391		67.1	16.22	1.273	4.190
3C84[b,c]	41.512		83.7	—	1.000	10.000
Venus 1991[d]						
Jan 01	−22.5	1.63496	32.3	29.380	1.00891	9.579
Jan 10	−20.2	1.61115	34.6	30.255	1.00918	9.862
Jan 20	−16.8	1.58115	38.0	31.414	1.00953	10.236
Jan 30	−12.6	1.54728	42.2	32.804	1.00995	10.685

[a] P. H. Richter, *Virgo-A Flux Density*, JPL IOM 3393-90-137 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 18, 1990.

[b] M. S. Gatti, M. J. Klein, and T. B. H. Kuiper, "32-GHz Performance of the DSS 14 70-Meter Antenna; 1989 Configuration;" *TDA Progress Report 42-99*, vol. July–September 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 206–219, November 15, 1989.

[c] Dummy value for $T100/C_r$ based on 9.628 found at f1.

[d] Declination of date

Table 7. Efficiency adjustments, deduced flux, and *T100/Cr* values: 32 GHz (Ka-band) at f3, January 1991

Source	Efficiency Adjustment Needed	Deduced Flux, Jy	Deduced $T100/C_r, K$
3C274[a]	1.03417	16.22	4.190
Venus[b]	0.96804	Table 6	Table 6
3C84[c]	1.02592	29.643	9.747

[a] Extended

[b] Variable

[c] Variable, point

Table 8. DSS BWG antenna X- and Ka-band final peak gains and efficiencies

	Gain, dBi	Efficiency, percent
8.45 GHz, X-band		
f1 pre-holography at 58.03 deg elevation	68.14	71.88
f1 post-holography at 45.63 deg elevation	68.34	75.35
f3 post-holography at 40.21 deg elevation	68.17	72.38
Estimated 1-σ errors	0.14 dB	2.3
32 GHz, Ka-band		
f1 post-holography at 44.82 deg elevation	78.33	52.31
f3 post-holography at 45.41 deg elevation	77.66	44.89
Estimated 1-σ errors	0.33 dB	3.6

Table 9. Coefficients of polynomial expressions for DSS 13 BWG antenna X- and Ka-band efficiencies without atmosphere

Efficiency	Efficiency, without atmosphere $= a_0 + a_1\theta + a_2\theta^2$ where θ = antenna elevation angle, degrees		
	a_0	a_1	a_2
X-band at f1	0.72734	0.00114701	−0.0000125694
X-band at f3	0.68496	0.00192912	−0.0000239880
Ka-band at f1	0.27220	0.01119613	−0.0001249108
Ka-band at f3	0.18285	0.01171770	−0.0001290241

Table 10. Final radio source flux and *T100/Cr* values deduced from DSS 13 34-m antenna-efficiency measurements

Frequency	Source	Flux, Jy	C_r	$T100/C_r$
8.45 GHz X-band	3C274[a,b]	44.555	1.087	13.477
	3C123[b]	—	—	3.092
	DR21[c]	—	—	6.523
	3C84[b,d]	32.67	1.000	10.629
	3C273[c,d]	29.30	1.000	9.635
32 GHz Ka-band	3C274[a,b]	16.22	1.273	4.190
	3C84[b,d]	29.462	1.000	9.688
	3C273[c]	31.418	1.000	10.331
	Venus[a,b,d]		—see text—	

[a] Calibration standard, no new values deduced

[b] f1 and f3 measurements

[c] f1 measurement only

[d] Variable source

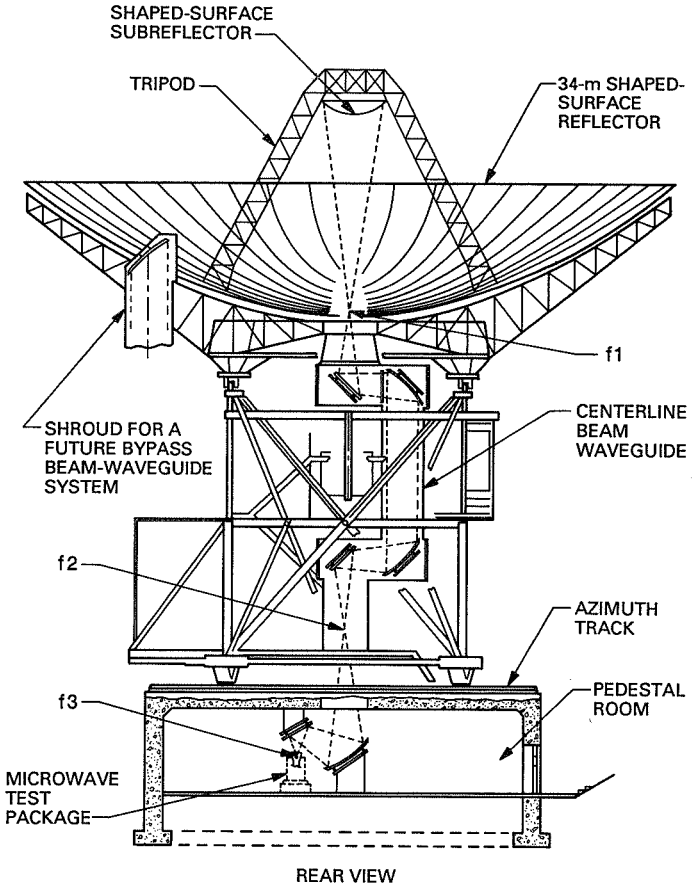


Fig. 1. Mechanical and microwave optics designs of NASA/JPL DSN beam waveguide antenna at Goldstone, DSS 13.

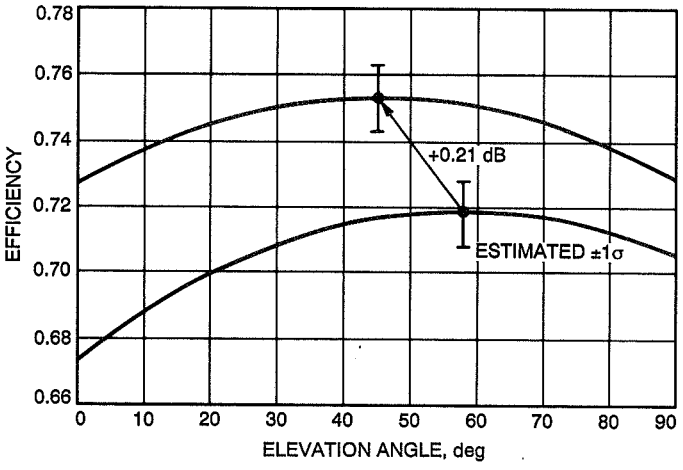


Fig. 2. DSS 13 8.45 GHz (X-band) efficiency at f1, pre-holography (bottom) and post-holography (top), without atmosphere.

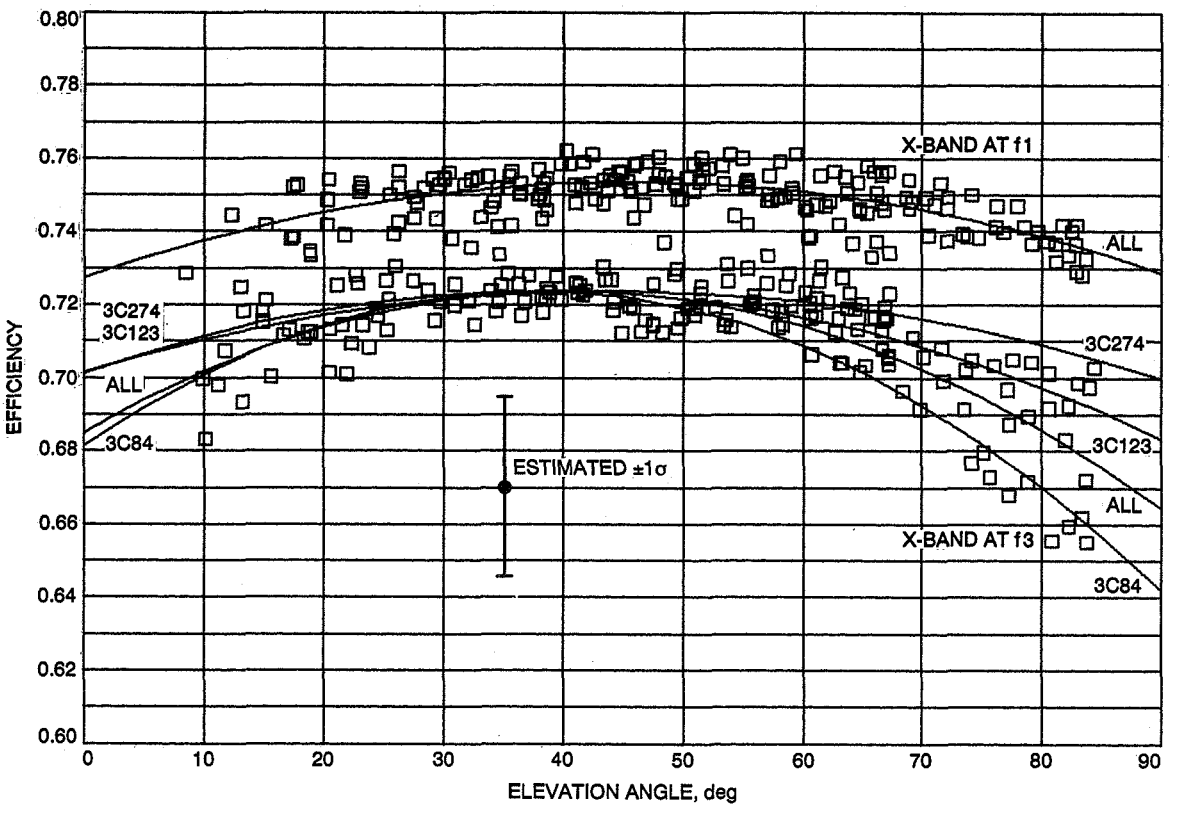


Fig. 3. DSS 13 8.45 GHz (X-band) efficiency at f1 and f3 focal points, without atmosphere.

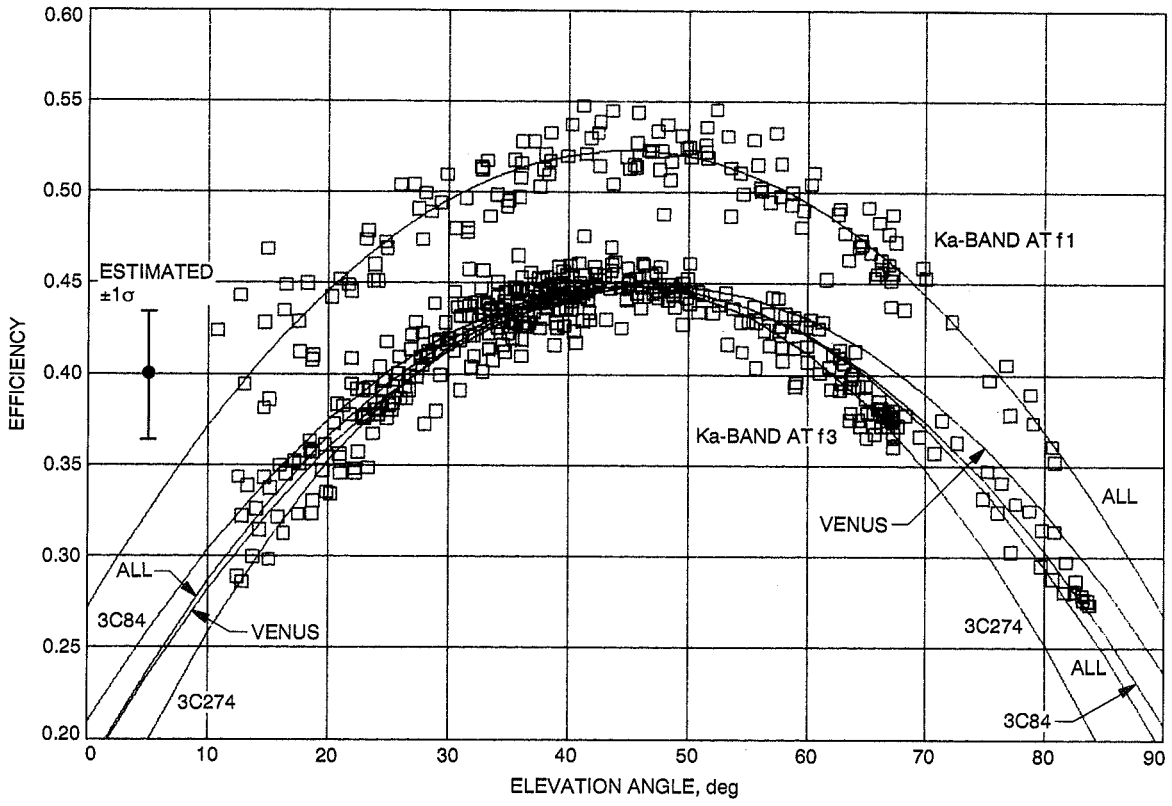


Fig. 4. DSS 13 32 GHz (Ka-band) efficiency at f1 and f3 focal points, without atmosphere.

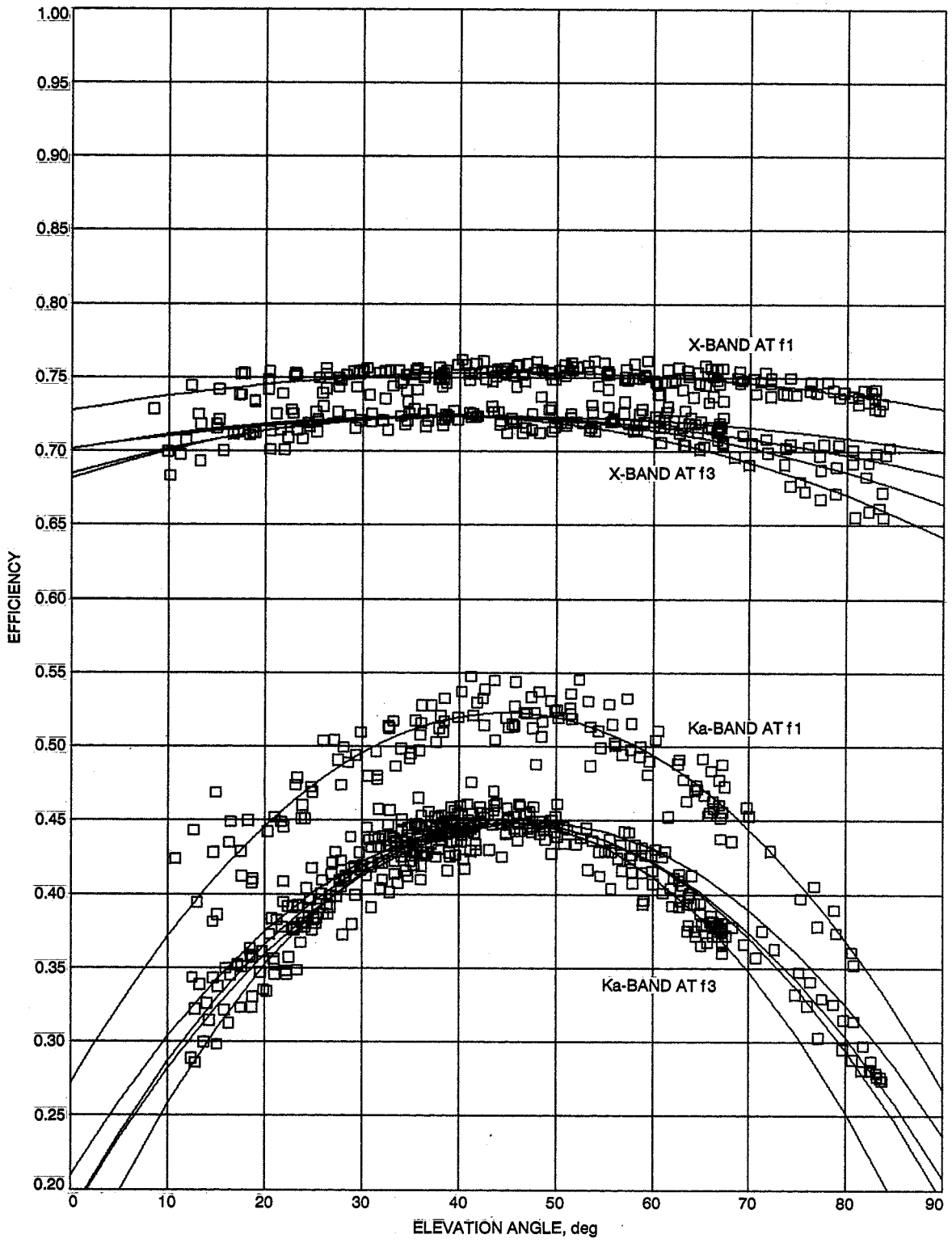


Fig. 5. DSS 13 X- and Ka-band efficiencies at f1 and f3 focal points, without atmosphere.

DSS 13 Phase II Pedestal Room Microwave Layout

T. Cwik and J. C. Chen
Ground Antennas and Facilities Engineering Section

This article describes the design and predicted performance of the microwave layout for three-band operation of the beam waveguide antenna DSS 13. Three pedestal-room microwave candidate layout designs were produced for simultaneous X-/S- and X-/Ka-band operation. One of the three designs was chosen based on given constraints, and for this design the microwave performance was estimated.

I. Introduction

This article describes the design and predicted performance of the microwave layout for three-band operation of the beam waveguide (BWG) antenna DSS 13 [1].[1] This design is part of the Phase II implementation of the antenna. The relevant microwave system is shown in Fig. 1 for a single feed in the pedestal room. The object of this work was to create pedestal-room microwave layouts for simultaneous X-/S- and X-/Ka-band operation. The center-band frequencies S-band (2.295 GHz), X-band (8.450 GHz), and Ka-band (32.000 GHz) were used in this design. Because the antenna had already been constructed, the obvious constraints of pedestal room size, BWG geometry, etc. were in place for this work. The size and shape of all BWG mirrors, including the basement ellipse, were set, although this mirror will be allowed to rotate to view different feed systems.

[1] M. Britcliff, ed., *DSS-13 Beam Waveguide Antenna Project: Phase I Final Report,* JPL D-8451 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 15, 1991.

II. Candidate Designs

Based on the above constraints, three candidate designs for the pedestal room were generated. The final design will take up some sector of the pedestal room floor, allowing for the possibility of the S-/X- and X-/Ka-band systems' having different optical paths, switched by the rotating ellipsoidal mirror. The defining constraint which emerged from the generation of these designs was the minimization of the area containing the apertures in the X-/Ka-band dichroic plate. (The dichroic plate has an inner aperture filled area along with an outer solid skirt.) This constraint will be examined in more detail in Section III.

A. Single-Sided, Vertical S-Band Design

The first design considered placed the S-, X-, and Ka-band horns vertically along one side of the ellipsoidal mirror (Fig. 2). S-/X- and X-/Ka-band dichroic plates and a Ka-band flat plate reflect the specific bands to the phase centers at f_3. The S-/X-band dichroic plate will be removed for X-/Ka-band operation since it is not designed

to pass the Ka-band. Vertical placement of the three horns preserves a single-plane 60-deg optical geometry. The immediate drawback of this design is the lack of room for a conventional S-band horn to be physically located in the existing pedestal room (Fig. 2 inset). The horn itself, without associated subsystems, extends approximately 66 in. below the existing pedestal room floor. Profiled S-band horns, which reduce the horn length, were also considered, but these did not eliminate the need for excavation of the pedestal room.

B. Single-Sided, Horizontal S-Band Horn Design

To eliminate the need to excavate the pedestal room, the second design places the S-band horn horizontally, as shown in Fig. 3(a). To preserve the correct electromagnetic optical design, the S-band horn and S-/X-band dichroic plate are both rotated in two planes. After considering the horn and plate to be fixed together, the proper orientation results from a 90-deg rotation up from the vertical to horizontal planes, and a rotation in the horizontal plane by 35.3 deg toward the ellipsoidal mirror. The S-band horn is now lying in the x-y plane at an angle such that rays from the horn reflect off the dichroic plate and down at a 60-deg angle to the ellipsoidal mirror. The plan view of this layout is shown in Fig. 3(b).

C. Double-Sided, Horizontal S-Band Design

The third candidate design is motivated by the close fit between feed systems and dichroic plates in the previous design. Depending on the feed support and associated amplifier and microwave subsystem, more space than shown is needed around the three feedhorns. For example, for an ultra-low noise amplifier Ka-band system, the dewar surrounding the Ka-band horn may extend 9 to 15 in. out from the center of the Ka-band horn, physically interfering with the X-/Ka-band dichroic plate. To gain more space, the double-sided design was produced, as shown in Fig. 4(a). The X-band feed system is duplicated, and the S-/X-band system is positioned along an optical path different from the X-/Ka-band system. This design relaxes the tight fit between feed systems and dichroic plates. The plan view of this arrangement is shown in Fig. 4(b). The 45-deg angle between the X-/S- and X-/Ka-band optical paths is chosen to allow clearance between the systems and to minimize the floor area used.

D. Computer-Aided Design of Candidate Designs

Because of the three-dimensional geometry, as well as the close fit between components in the designs, a computer-aided design (CAD) system was used to draft each of the three candidates. The ellipsoidal mirror was drawn into the pedestal room, and the various components added. Estimated thicknesses of the flat and dichroic plates were included as well as the backing of the ellipsoidal mirror. Feedhorn mounting fixtures and the envelopes of anticipated microwave subassemblies were similarly added and color-coded by frequency band. The three candidate designs were drawn separately and optic rays were added to define the optical envelope of the layout. The CAD system can continuously rotate and zoom the image on the screen, allowing visualization of the design from different viewing angles and distances, and has the capability of measuring distances among components. To provide a hard copy of this visualization, a video recorder was used to create an 8-min video of the three designs as the images were rotated and zoomed.[2] This video is available from the authors for viewing.

III. X-/Ka-Band Dichroic Plate Constraint

The dichroic plates needed in the above layouts are for simultaneous Ka-band (pass band) and X-band (reflecting band), and X-band (pass band) and S-band (reflecting band) operations. It was found that the X-/Ka-band plate was a defining constraint of the layout.

The location and the size of the dichroic plates in the layouts are dependent on the spillover loss, mechanical layout, cost, and the constraints of fabrication. The dichroic plates need to meet three criteria: (1) The size is electromagnetically big enough to minimize the spillover loss—the bigger the dichroic plate, the lower the spillover loss at a given location. (2) The closer the plate is to the focal point of the ellipsoidal mirror, the smaller the spillover loss is for a given-size plate. Hence, the dichroic plate is placed as close to the focal point of the ellipsoid as possible to reduce the plate size and minimize spillover loss, but to give enough clearance to other components in the mechanical layout. (3) The cost of the dichroic plate is roughly proportional to the number of holes on the plate. Therefore, the size of the dichroic plate has to remain within the budget and under a 38-in. diameter, which is the maximum size that can be fabricated.

The layout for the X-/Ka-band system is shown in Fig. 4(a). The X-/Ka-band dichroic plate is a 28-in.-diameter round plate with approximately 12,000 holes and is 28 in. away from the focal point of the ellipsoidal mirror in the pedestal room. The mechanical layout shows no

[2] T. Cwik, J. Chen, and G. Hale, "DSS 13 Phase II Microwave Layout Designs," videotape, Ground Antennas and Facilities Engineering Section, Jet Propulsion Laboratory, Pasadena, California, August 1990.

blockage and clearance to all nearby components, giving the minimum size for the X-/Ka-band plate.

IV. Design Choice

Based on the above constraints and layouts, the third candidate was chosen for implementation. The first candidate was ruled out because of the need for excavating the pedestal room. Of the second and third candidates, the third was chosen because of maximum room among components. If floor space becomes a premium, the components can be rearranged into the second design, allowing for future systems to be implemented.

V. Predicted Performance

The gain of the antenna for the chosen design layout was estimated by physical optics (PO) integration of currents induced on the mirrors. The breakdown of gain and spillover losses in the beam waveguide and a benchmark calculation at f_1 are shown in Fig. 5 for the three bands. The estimate for a given frequency is found from a succession of PO integrations, beginning with a calculation of the fields reflected from the ellipse due to a feed pattern at f_3. This field distribution is the input pattern used in a PO–PO integration through the two parabolas P_1 and P_2. Flat mirrors F_1 and F_2 are neglected in this calculation.

The field distribution emerging from P_1 is then the input to a PO/Jacobi-Bessel calculation for the field radiated from the sub-/main reflector system. The listed gain at f_3 is the result of this series of calculations. Spillover estimates past the BWG mirrors are also listed in Fig. 5. Spillover past the basement mirrors (B) is the amount of energy spilling past the mirrors in the electromagnetic path at the given frequency. For example, at the X-/Ka-band side, the Ka-band feed distribution sees the Ka-band flat plate and X-/Ka-band dichroic plate, while the X-band field distribution sees only the X-/Ka-band dichroic plate. Spillover past the flat plates F_1 and F_2 was estimated indirectly from a separate calculation, rather than from the succession of PO calculations. For a benchmark estimate, the antenna gain due to a 29-dBi horn pattern referenced at f_1 is included.

Extra analysis was extended to the S-band system because of the loss of gain relative to a theoretical feed pattern at f_1. The central conclusion drawn was that the S-band field defocuses as it propagates through the BWG. That is, for a 22.4-dBi S-band horn referenced to f_3, the resulting field propagated through the BWG and referenced to f_1 is not a 29-dBi amplitude field with corresponding phase distribution that properly illuminates the sub-/main reflector system, producing maximum gain. The defocusing mainly results as the feed pattern is magnified by the ellipsoidal basement mirror. Ideally, the 22.4-dBi feed pattern at f_3 should refocus into a 29-dBi pattern referenced to f_2. This field would then propagate through the next BWG section to properly refocus at f_1 with no change of gain. It was found that the S-band field refocuses approximately 90 in. above the f_2 focus, improperly illuminating the next section of the BWG.

Numerical experiments were made to move the focus back toward f_2. A method was developed which partially accomplished this by defocusing the horn at f_3, while changing its gain. For example, by defocusing the horn away from the S-/X-band dichroic plate and increasing the horn gain, the resultant focus could be pulled back toward f_2. Although this technique could modify the horn focus relative to f_2, the change in overall antenna gain was not significant enough to add the defocusing to the design. The conclusion drawn from the S-band study was that the ellipsoidal mirror, being approximately 20 wavelengths at S-band, is simply not large enough to focus the field near f_2 and therefore properly illuminate the next section of the BWG.

VI. Epilogue

Implementation of this design began in the first quarter of 1991. Further Ka-band design work is being considered due to dichroic plate cost and size of the ultra-low noise amplifier and associated dewar.

Acknowledgment

The authors gratefully acknowledge the skilled assistance of Gary Hale of the Prototype and Mechanical Fabrication Services Section, for the CAD of the above layouts.

Reference

[1] T. Veruttipong, W. Imbriale, and D. Bathker, "Design and Performance Analysis of the DSS-13 Beam Waveguide Antenna," *TDA Progress Report 42-101,* vol. January–March 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 99–113, May 15, 1990.

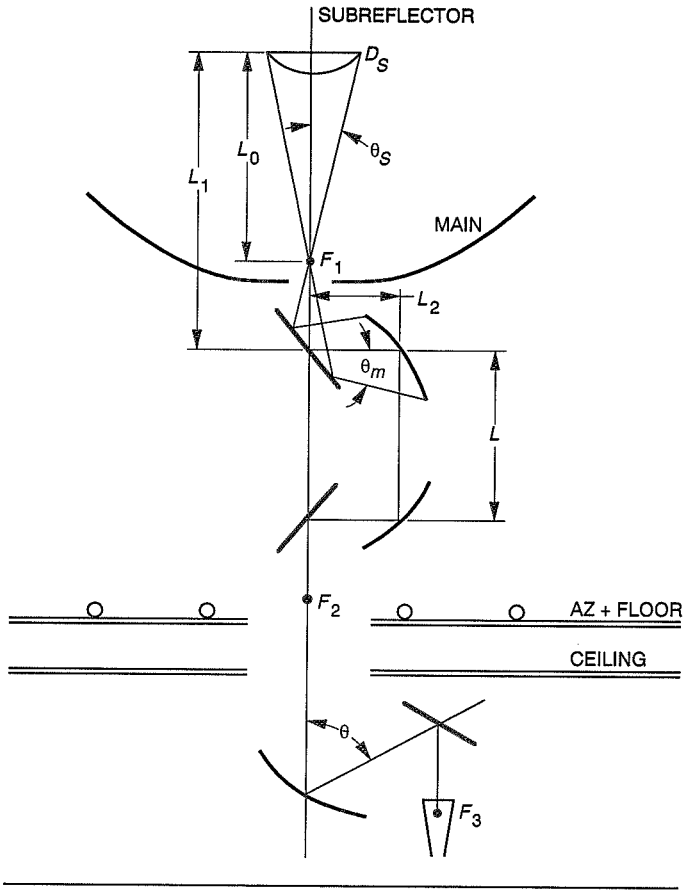


Fig. 1. Beam waveguide optics of DSS 13.

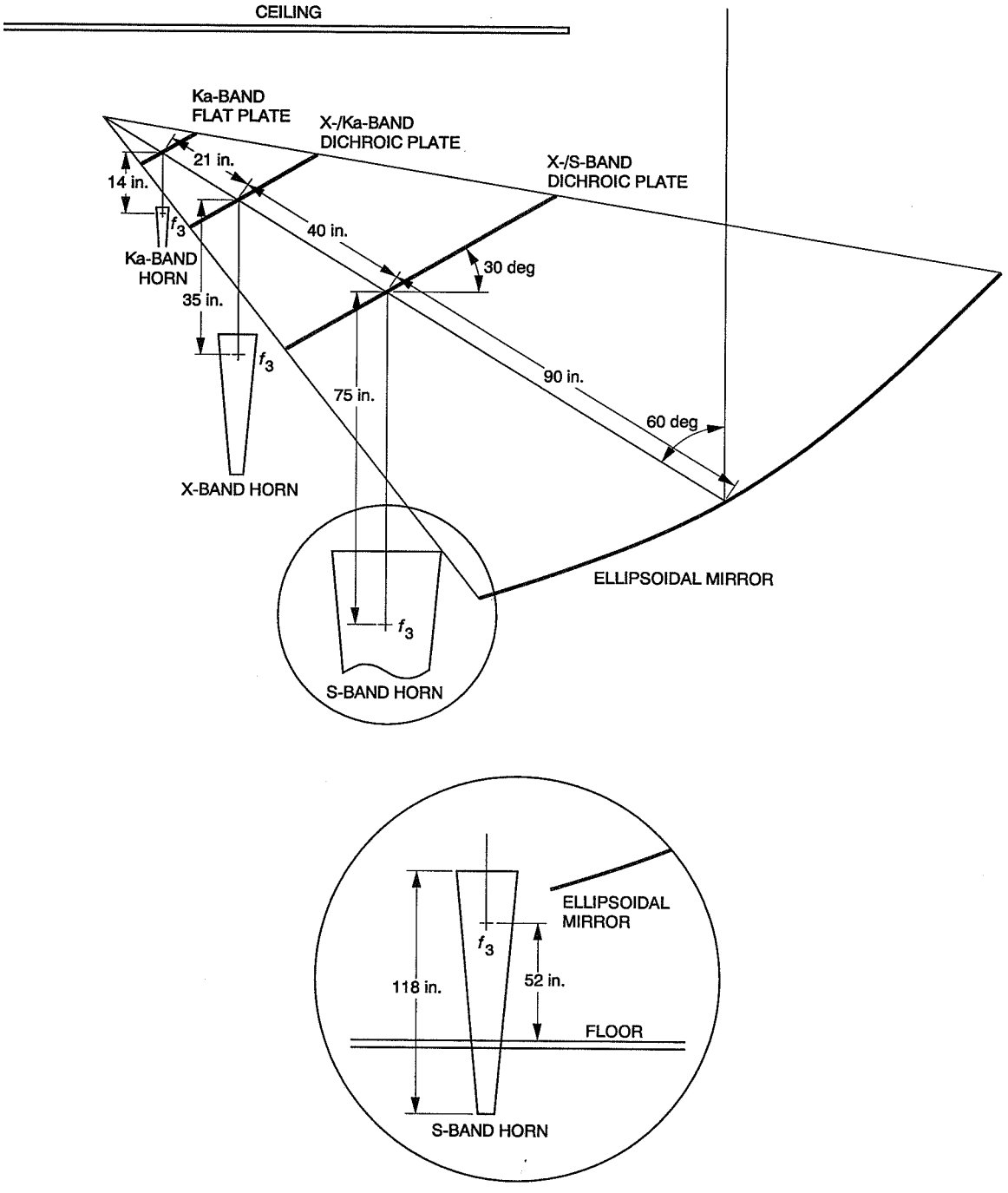


Fig. 2. Pedestal room optics for single-sided, vertical S-band horn design. Inset shows detail of vertical S-band horn with existing floor.

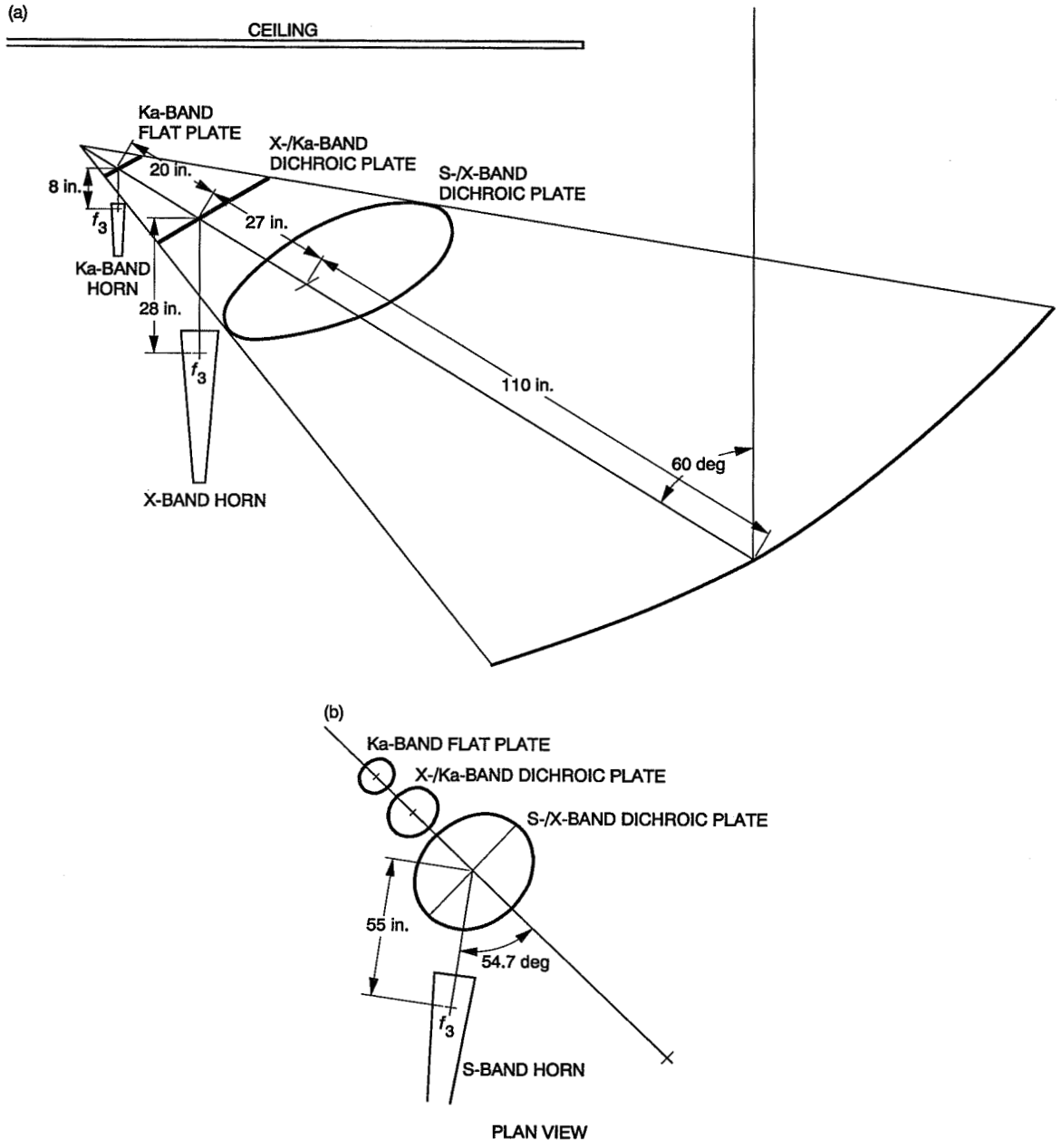


Fig. 3. Single-sided horizontal S-band horn geometry: (a) pedestal room optics and (b) plan view.

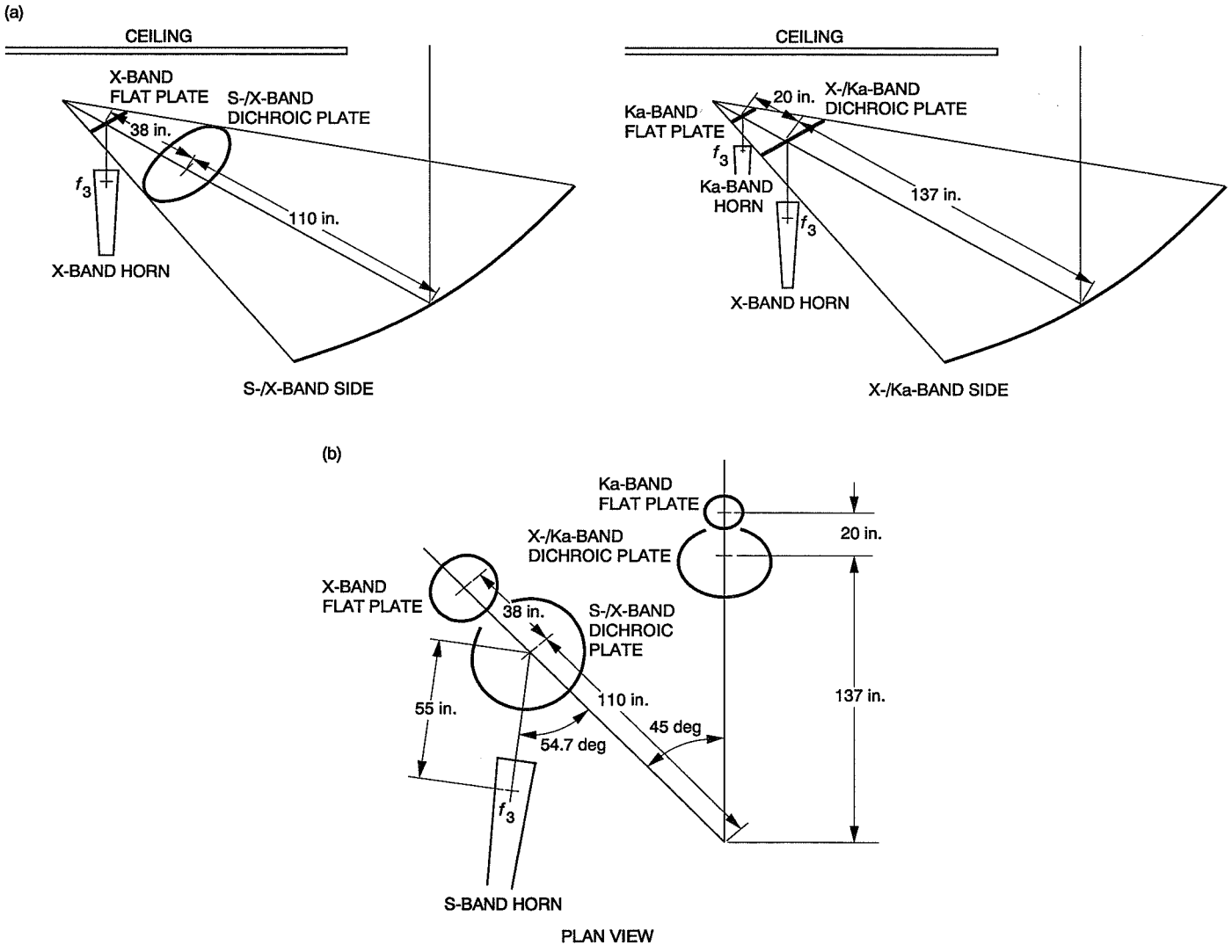
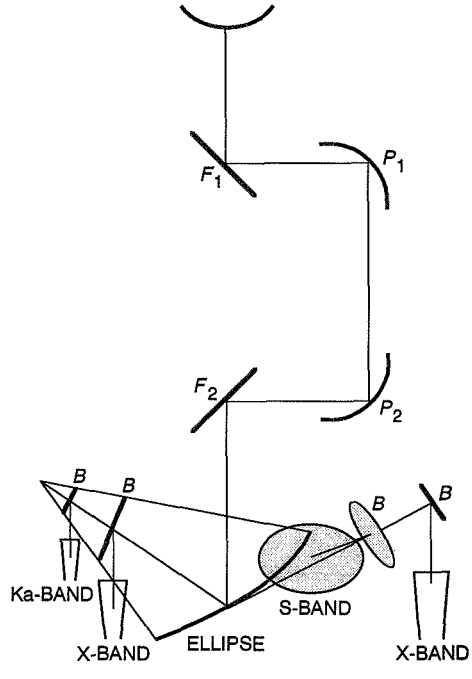


Fig. 4. Double-sided horizontal S-band horn geometry: (a) S-/X- and X-/Ka-band and (b) plan view with 45-deg offset angle.



BAND	BWG SPILLOVER, dBi					GAIN[a], dBi f_3	GAIN[b], dBi f_1
	ΣB	ELLIPSE	$P_1 + P_2$	$F_1 + F_2$	TOTAL		
S	0.01	0.08	0.55	0.14	0.78	55.85	57.47
X	0.03	0.07	0.06	0.02	0.17	69.00	69.20
Ka	0.03	0.06	0.05	0.01	0.14	80.54	80.70

[a] DOES NOT INCLUDE RMS, DICHROIC, WAVEGUIDE, etc. LOSSES.
[b] 29-dBi HORN PATTERN REFERENCED AT f_1.

Fig. 5. Beam waveguide geometry and predicted performance for double-sided design.

A Carrier-Arraying Demonstration at Goldstone for Receiving Pioneer 11 Signals

T. T. Pham, M. K. Simon, and T. K. Peng
Telecommunications Systems Section

M. H. Brockman[1]

S. S. Kent[2]

R. Weller[3]

A carrier-arraying technique was demonstrated at Goldstone in May 1990. The Block III receivers of two 34-m antennas, DSS 12 and DSS 15, were arrayed together to receive S-band (2.3-GHz) signals from the Pioneer 11 spacecraft. Carrier phases in the two receivers were synchronized by the analog phase-lock loops, and carrier signals were added at an intermediate frequency to enhance tracking performance. The receiver at DSS 15, which had been unable to lock up and track the Pioneer 11 signal by itself due to a wider tracking loop bandwidth and a higher system temperature, was now able to track the carrier and produce usable baseband signals. The receiver at DSS 12 achieved a reduction of the rms phase error, increasing the telemetry symbol SNR by an average of 0.35 dB. The baseband signals from both antennas were then synchronized and combined using the existing Baseband Assembly, thereby achieving a total symbol SNR increase of 2.5 ± 0.7 dB relative to DSS 12 alone. Baseband combining would have been impossible without carrier arraying in this case. In analyzing the performance of carrier arraying, previous models only treated the rms phase errors caused by the ground receiver thermal noises. But as very narrow tracking loops were employed, the contribution of phase noises transmitted from the spacecraft or generated in the ground-receiver oscillators should be included. In this article, a more comprehensive model is presented to permit evaluation of both thermal and phase-noise effects. The analysis agrees well with observed data.

[1] Consultant, M. H. Brockman and Associates.

[2] Consultant, Mainstream Engineering.

[3] CDI Consultants.

I. Introduction

The Deep Space Network (DSN) has the capability to combine baseband signals from several antennas by using the Baseband Assembly (BBA). Baseband combining depends on baseband signals to be demodulated at each antenna, which requires the receiver at each antenna to be in lock. Some distant spacecraft, such as Pioneer 10 and 11, have signal levels so low that only the 70-m antenna with the narrowest (3-Hz bandwidth) receiver tracking loop can track these signals with adequate performance, as shown in Table 1. As of May 1990, the 34-m standard (STD) antenna could receive these signals using a 3-Hz loop, with unsteady performance. However, the 34-m high-efficiency antenna (high efficiency at 8.4 GHz but not at 2.3 GHz) was unable to acquire or track the signals at all because of its higher system temperature and wider loop bandwidth.

The technique described in this article, as depicted conceptually in Fig. 1, enabled both of the 34-m antennas to lock onto the weak carrier signals and produce baseband signals in an engineering demonstration conducted at Goldstone, California. The local oscillator signal of DSS 12, a 34-m STD antenna station, was sent to DSS 15, a 34-m high-efficiency (HEF) antenna station, to aid carrier acquisition and tracking there. Since the Doppler frequency difference between the two stations was small (less than ± 6 Hz), a very narrow loop could be used at DSS 15 to track the carrier. The carrier signal from DSS 15 was then sent at an intermediate frequency (IF) to DSS 12 to be combined with the carrier signal at DSS 12. Receiver performance at both antennas was improved.

The general term *carrier arraying* is used to refer to both carrier-aiding and carrier-combining operations. In this demonstration, carrier aiding enabled DSS 15 to acquire and track the spacecraft signal. Carrier combining further improved tracking performance at DSS 12 and DSS 15.

This carrier-arraying technique was first suggested in [1]. The performance was analyzed in detail in [2], including the effects of thermal noises from the ground stations. Specific design parameters for antennas with different gain/temperature (G/T) ratios were suggested in [3]. Hardware modules implementing this technique were first installed in the receiver development laboratory at JPL and tested in January 1990. The same modules were then installed at DSS 12 and DSS 15 in April 1990 for field testing. Pioneer 11 telemetry signals were successfully received by the system on May 3 and 6, 1990.

Section II describes the configuration of the carrier array system used in the demonstration. Section III describes the data received from the spacecraft and discusses the observed effects of carrier arraying on receiver performance. Section IV includes a comparison of actual data with the calculation of an enhanced mathematical model. This model, given in detail in Appendix A, permits evaluation of both thermal noise and phase-noise effects. Section V includes the conclusions. Appendix B includes an analysis of the optimal choice of weighting factors for carrier combining. Also given are the equations from the existing phase-lock-loop and carrier-arraying theories relevant to this study.

II. System Description

A. Configuration

A detailed block diagram of the modified Block III receivers at DSS 12 and DSS 15 is given in Fig. 2. The modification includes modified or new hardware modules for carrier arraying and three fiber-optic lines connecting the two receivers: the 70-MHz local oscillator (LO) from DSS 12 to DSS 15, the 10-MHz IF signal from DSS 15 to DSS 12, and the 100-MHz frequency reference from DSS 15 to DSS 12. A series of switches was used to allow the receivers to operate in either "normal" or "carrier array" mode, identified as "N" or "A" in Fig. 2. Normal mode permits the receiver to provide regular flight support without using the array modules.

During the demonstration, the first LO into the S-band mixer at DSS 15 was always aided by the DSS 12 LO signal. The DSS 15 carrier was combined with the DSS 12 carrier at 10-MHz IF by a switch at DSS 12. The combiner was switched in and out several times to measure the effect of carrier combining.

At DSS 12, the downlink signal was first amplified by the maser low-noise amplifier, with a system noise temperature around 20 K at zenith. The signal was downconverted with a reference signal coming from the carrier tracking loop. The output at 50 MHz was further downconverted to 10 MHz. One of the 10-MHz IF distribution amplifier outputs was used for baseband demodulation. The other 10-MHz IF signal passed through a 550-Hz predetection filter into the carrier tracking loop. This filtered 10-MHz IF signal could be combined with a similar signal from DSS 15 by the action of the carrier combiner switch. The power levels of the two combining carriers were kept relatively constant by the automatic gain control (AGC) in the Block III receivers. The 70-MHz LO signal from this carrier tracking loop was distributed to DSS 15 to drive its first local oscillator.

At DSS 15, the received S-band signal was first amplified by the high-electron-mobility transistor (HEMT) low-noise amplifier with a noise temperature around 35 K at zenith. It was then mixed with the reference signal originating from DSS 12, producing a 50-MHz IF at the S-band mixer output. The 50-MHz signal was again mixed with a 60-MHz signal from the phase-locked loop to produce a 10-MHz IF. This IF signal was further converted to baseband. A separate branch of this signal was filtered through a 4950-Hz predetection filter, and subsequently a much narrower filter, for carrier tracking. The 10-MHz IF signal after the 4950-Hz filter was also distributed to DSS 12 for carrier combining.

Primary modifications to the existing Block III receivers for carrier arraying included the following:

1. The 3-Hz Tracking Loop at DSS 12. The carrier tracking loop had a threshold loop bandwidth ($2B_{LO}$) of 3 Hz and a loop damping factor (r_0) of 4. This loop bandwidth was narrow enough to provide sufficient carrier signal-to-noise ratio in the loop and large enough to track the carrier in the presence of Doppler uncertainty without aiding.

2. The 0.13-Hz Tracking Loop at DSS 15. Since the operating noise temperature at DSS 15 was higher than that at DSS 12, a narrower tracking loop with a threshold loop bandwidth ($2B_{LO}$) of 0.13 Hz was used to reduce the phase error due to thermal noise. Such a narrow loop was adequate to track the carrier signal because it only had to track the Doppler frequency difference between the two antennas. Most of the Doppler frequency shift due to Earth rotation was tracked out by the 3-Hz loop of the DSS 12 receiver through the common 70-MHz LO.

3. Common Frequency Reference Between DSS 12 and DSS 15. The receivers at DSS 12 and DSS 15 shared the same source of reference frequency through a fiber-optic link. This common reference eliminated any potential phase drift between the two IF signals to be combined and any apparent Doppler in the DSS 15 receiver due to a drifting 70-MHz LO.

4. Gain Modification To Reduce the Interference of Spurious Signals. When an initial attempt was made at DSS 15 to acquire the low-level Pioneer carrier signals, spurious signals of significant power relative to the carrier were detected. One spurious signal adjacent to the 50-MHz IF interfered with the signal acquisition. A 20-dB amplifier had to be installed upstream from the mixer to reduce the spurious signal effect to an acceptable level. As carrier arraying is only needed when the signal level is near the thermal noise threshold, low-level spurious signals that have no adverse effect on other occasions can actually interfere with the signal reception in this case. This experience highlighted the need for careful testing of spurious signals before committing the new system to flight support.

B. Operational Procedure

Prior to the tracking, the telemetry detector phase was adjusted to remove any bias. The carrier combiner was adjusted to align the phases of the two signals and set the weighting factors for combining. The factors were set to unity in this demonstration. Later analysis, given in Appendix B, indicated that a different choice of weighting factors could have been used to optimize the signal-to-noise ratio in the tracking loop at DSS 12. That choice would have increased the carrier tracking loop SNR (P_c/N_0B_L) by 0.13 dB.

Because of the narrow tracking loop employed, it was not possible for DSS 15 to acquire the signal independently. The acquisition process always started with the receiver at DSS 12. After phase lock, DSS 12 sent its first LO to DSS 15. The DSS 15 receiver was then tuned to compensate for the frequency difference between the two stations, allowing the DSS 15 receiver to acquire the signal. Carrier combining was ready to be activated after both stations established lock.

III. Data Received From the Spacecraft

A. Results

1. May 3, 1990. Figure 3 presents the symbol SNR (ST_s/N_0) of the carrier-arrayed and baseband-combined telemetry measured on May 3. The telemetry rate was 16 bits per sec convolutionally coded, or 32 symbols per sec modulated on a 32-kHz subcarrier. Each data point in Fig. 3, and later in Figs. 4(a) and 5, represents measurements at the interval of 60 sec. At all times, the DSS 15 receiver was carrier-aided from DSS 12. Carrier combining was activated in periods 8:45–9:30 GMT and 10:00–10:15 GMT and was deactivated from 9:30–10:00 GMT. Only the performance of the DSS 12/DSS 15 baseband-combined telemetry was directly measured. DSS 12 or DSS 15 telemetry symbol SNR could not be individually measured due to equipment constraints. During this period, the system operated in a 3-way mode with uplink signals transmitted from DSS 61, a 34-m antenna in Madrid, Spain. The system noise temperature at zenith

was recorded as 21.2 K at DSS 12 and 35.4 K at DSS 15. However, the effective noise temperature was 3 K higher since the signals were received in the elevation range of 42–46 deg.

On this date, the DSS 14 station also tracked Pioneer 11 spacecraft and its measured symbol SNR is presented in Fig. 4(a). A drop of 0.3 dB was observed from 1-way to 3-way operation.

Normally the received carrier signal level is measured with the AGC signal provided by the receiver. This measurement was not made at DSS 12 because the receiver gain was changed and was not calibrated due to lack of time. Instead, the carrier signal level at DSS 12 was inferred from the -166.2 ± 0.3-dBm level measured at DSS 14, as shown in Fig. 4(b). Since the receiving antenna gains are known to differ by 7.3 ± 0.3 dB,[4] the carrier power at DSS 12 was estimated to be -173.5 ± 0.4 dBm. All uncertainties are referenced to 1σ.

2. May 6, 1990. A more comprehensive set of measurements was made, as shown in Fig. 5. The DSS 12 telemetry and the baseband-combined DSS 12/DSS 15 telemetry were individually measured, and 1-way and 3-way telemetry data were both available. Carrier combining was activated and deactivated many times to observe its effects.

The system noise temperatures at zenith were 25.6 K at DSS 12 and 36.5 K at DSS 15 (an increase in the noise temperature at DSS 12 was a result of the microwave switches having been configured in a diplexed mode, as compared with the nondiplexed configuration on May 3). For the elevation range in which data were taken, i.e., 21–47 deg, the effective noise temperature at DSS 12 was 32.1–28.6 K at DSS 12 and 43.0–39.5 K at DSS 15. In the 1-way mode, the track started with receiver carriers combined from 6:45–8:15 GMT. Carrier combining was turned off in the next 15 minutes from 8:15–8:30. At 8:35, the spacecraft signals changed to 3-way mode. In the next one and a half hours, the carrier-combining option was again activated twice from 9:00–9:45 and 10:00–10:15.

Note that the symbol SNR dropped by 0.2 to 0.3 dB when the spacecraft signal changed from 1-way to 3-way mode as a result of transponder uplink noise. Note also that the baseband-combined data in the 3-way mode on May 6 were about 0.9 dB lower than the data on May 3. This was caused by the difference in the system noise temperature (0.8-dB difference at 40-deg elevation) on two different passes.

3. Statistical Analysis of Received Data. In analyzing the data, a least-squared-fitting algorithm that included the known elevation-angle dependence on antenna gain and system temperature was used to generate best-estimate curves. Any measurement of telemetry symbol SNR was modeled as the input symbol SNR minus a degradation in the receiver, minus another degradation in the subcarrier demodulation and symbol synchronization process, plus a baseband-combining gain, if appropriate. The input symbol SNR was determined from the measured carrier power, a known modulation index, and a known functional dependence on elevation angle. The degradation in subcarrier demodulation and symbol synchronization was assumed to be 0.8 dB when the Subcarrier Demodulator Assembly/Symbol Synchronizer Assembly (SDA/SSA) was used and 0.3 dB when the BBA was used.[5] The baseband-combining gain reflected the algebraic sum of the symbol SNRs of the two baseband signals, assuming that the two signals were optimally combined [4]. The algorithm solved the receiver degradation as an unknown constant in the least-squared sense, thus defining the curve.

Only data collected under the same configuration (1-way versus 3-way, carrier-combined versus noncombined, single-station versus baseband-combined) were fitted together as an ensemble. The results of these fittings are the curves given in Figs. 3, 4(a), and 5.

B. Discussions

1. Effect of Carrier Arraying on DSS 15 Performance. The 70-MHz local oscillator from DSS 12 enabled the DSS 15 receiver to acquire and track the carrier signal and demodulate the telemetry baseband. Symbol SNR of the DSS 15 telemetry was not measured directly. However, it could be inferred from the DSS 12 telemetry symbol SNR and the known difference in system temperature. After using this inferred value to calculate the symbol SNR of the baseband-combined telemetry of DSS 12 and DSS 15, the estimate agreed with the measurement to within 0.2 dB.

2. Effect of Carrier Combining on the DSS 12 Telemetry. In the stand-alone mode, DSS 12 performed very close to the prediction. For example, the measured symbol SNR was 0.1 dB at 8:15 GMT on May 6 (Fig. 5),

[4] *Deep Space Network/Flight Project Interface Design Handbook*, 810-5, Rev. D, vol. 1, modules TCI-10 and TCI-30 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 1, 1990.

[5] Ibid., module TLM-30.

while the predicted value was 0.3 dB based on G/T and the estimated symbol SNR degradation in the system. When the carriers were combined, the symbol SNR at DSS 12 increased by 0.35±0.5 dB (Fig. 5 or Table 4, 3-way mode). This is consistent with the estimated 0.4-dB improvement based on a model given in Section IV below.

3. Effect of Carrier Arraying and Baseband Combining Together. Carrier arraying enabled DSS 15 to lock up the receiver and produce telemetry baseband signals. Baseband combining added DSS 12 and DSS 15 signals together, resulting in an observed 2.15 ± 0.5-dB increase in symbol SNR relative to DSS 12 telemetry (Table 4, 1-way). Given the G/T ratios of both antennas and the estimated comparable degradations in both systems, this baseband-combining gain is consistent with expectations.

4. Telemetry Degradation When the Spacecraft Downlink Changes From 1-Way to 3-Way Mode. The 0.2–0.3-dB degradation observed in Fig. 5 is not surprising since the phase noise generated on the spacecraft downlink due to thermal noise on the uplink contributed an additional amount of phase error in the ground receiver carrier tracking loop. This phenomenon, which was also observed independently at DSS 14, is explained in [5].

IV. Performance of Data Compared With Analytical Model

A. An Improved Mathematical Model for Carrier Arraying

Since telemetry degradation in the carrier demodulation process depends on the mean-squared value of the total phase error that appears at the output of the phase detector, every source of carrier phase error contributes to telemetry degradation. Existing mathematical models only address the contribution of the thermal noise referring to the front end of the ground receiver [2]. Yet, it is known that phase noises contained in the received spacecraft signal and phase noises generated in the ground receiver (normally in the voltage-controlled oscillator [VCO]) also contribute to the phase error. In the months following the demonstration, the mathematical model given in [2] was extended to take into account phase noises generated on the spacecraft and in the receiver VCO. This extended model is described in detail in Appendix A. Application of the model to the specific array environments at DSS 12 and DSS 15 is explained below.

1. Mean-Squared Phase Error in the First Receiver (DSS 12). For a two-receiver system, the mean-squared phase error observed in the main receiver (DSS 12) in an array configuration can be expressed as

$$\sigma^2_{\phi_1} = \frac{1}{2\pi j}\int \left|\frac{H_1(s)}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 ds \frac{N_{01}}{2P_1\cos^2\theta_m} + \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1-H_2(s)]}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 ds \frac{N_{02}}{2P_2\cos^2\theta_m}$$

$$+ \frac{1}{2\pi j}\int \left|\frac{1-H_1(s)}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 S_{\Delta\psi 1}(s)ds + \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1-H_2(s)]}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 S_{\psi R2}(s)ds \quad (1)$$

where

$H_1(s)$ = the closed-loop transfer function of the DSS 12 receiver

$H_2(s)$ = the closed-loop transfer function of the DSS 15 receiver

P_1, P_2 = the total received power at DSS 12 and DSS 15, respectively

N_{01}, N_{02} = thermal noise spectral densities of DSS 12 and DSS 15

θ_m = the telemetry modulation index

$S_{\Delta\psi 1}$ = the noise spectral density of the spacecraft phase noise and the DSS 12 receiver VCO phase noise (independent sources)

$S_{\psi R2}$ = the noise spectral density of the DSS 15 receiver phase noise

This equation is taken from Eq. (A-15) of Appendix A with two antennas ($N = 2$), unity carrier combining weighting factor ($\beta = 1$), and the same amplitude of car-

rier signals ($\gamma = 1$). As a result, the carrier-array gain G, as defined in Eq. (A-16), equals 2 when the carriers are combined and 1 when the carriers are not combined.

The first term in Eq. (1) is the contribution of thermal noise from DSS 12. The second term is the thermal noise contribution from the DSS 15 receiver to the DSS 12 receiver through carrier combining at IF. Spacecraft phase noise and DSS 12 VCO phase-noise contribution are seen in the third term. The fourth term describes the contribution of the DSS 15 receiver phase noise, which affects DSS 12 through carrier combining.

Notice that the effects of the spacecraft and VCO phase noise are filtered with a high-pass function $[1 - H(s)]$, whereas thermal noise is filtered through a low-pass function $H(s)$. Since the phase-noise spectral density at the VCO output is often modeled as $1/f^3$, careful consideration needs to be given to the choice of loop bandwidth. The lower the loop bandwidth, the more contribution can be expected from the VCO. If the VCO or spacecraft phase noise dominates, a tracking loop of larger bandwidth is better. On the other hand, a narrower loop is desirable if thermal noise becomes dominant.

When the 10-MHz IF from DSS 15 is not combined with that of DSS 12, the second and fourth terms become zero and the factor G becomes unity. The equation then reduces to the well-known phase error expression for a single receiver.

2. Mean-Squared Phase Error in the Second Receiver (DSS 15). Similarly, the phase error in the second receiver at DSS 15 is summarized in the following equation:

$$\sigma^2_{\phi_2} = \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1-H_2(s)]}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 ds \frac{N_{01}}{2P_1\cos^2\theta_m} + \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1-H_2(s)](G-1)+H_2(s)}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 ds \frac{N_{02}}{2P_2\cos^2\theta_m}$$

$$+ \frac{1}{2\pi j}\int \left|\frac{[1-H_1(s)][1-H_2(s)]}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 S_{\Delta\psi 1}(s)\,ds + \frac{1}{2\pi j}\int \left|\frac{[1-H_2(s)]}{1+H_1(s)[1-H_2(s)](G-1)}\right|^2 S_{\psi R2}(s)\,ds \tag{2}$$

where all parameters are defined as in Eq. (1).

This equation is taken from Eq. (A-25) in Appendix A. The first term is the thermal noise coming from the DSS 12 receiver through the first LO. The second term consists of two parts, both caused by thermal noise in the DSS 15 receiver itself. The first part, $H_1(s)[1 - H_2(s)]$, passes the receiver thermal noise from DSS 15 to DSS 12 via carrier combiner and then back to the DSS 15 through the first LO. The second portion, $H_2(s)$, is the well-known effect of the thermal noise from the DSS 15 front end to the DSS 15 receiver itself. The third term represents the phase-noise contribution from the spacecraft and DSS 12 VCO through the first LO. The DSS 15 receiver VCO phase-noise contribution is shown in the last term.

B. Predicted Phase Errors in the 1-Way Mode

This model was used to estimate the mean-squared phase errors in both receivers for this demonstration. The phase-noise spectral density from the spacecraft oscillator is assumed to vary as c/f^3, where f is the frequency offset from the carrier, and c is a coefficient that depends on the phase-noise characteristics of the VCO and the frequency multiplier following it. The value of c for Pioneer 10, a similar spacecraft, was previously determined from measurements to be approximately $4\pi \times 10^{-4}$ [6]. The coefficient for the Block III receiver VCO, with a frequency multiplier of 96 as used at DSS 12, was determined to be 4.4×10^{-3} in [7]. The coefficient for the DSS 15 receiver was much smaller because of a smaller frequency multiplier of 3, instead of 96, used for the carrier-arraying tracking loop. The value of c for the receiver at DSS 15 was therefore reduced by a factor of $(3/96)^2$ to 4.3×10^{-6}. The carrier-to-noise density ratio (P_c/N_0) used in the calculation reflected the actual condition during the demonstration. The calculated mean-squared phase errors contributed by all sources at the DSS 12 receiver are given in Table 2.

Several comments can be offered. First, the thermal noise effect of the DSS 12 receiver was reduced in the

carrier-combined mode because of the presence of a limiter. Specifically, as the 10-MHz IF signal from DSS 15 was injected into the carrier combiner at DSS 12, noise of the 4950-Hz bandwidth from DSS 15 came with it. This resulted in a higher suppression of signal power and, in effect, the loop gain and loop bandwidth were reduced, as reflected in Eqs. (A-18) and (A-19). Secondly, the transfer function $H_1(f)[1 - H_2(f)]$ for the thermal noise of the DSS 15 receiver has effectively the same bandwidth as that of $H_1(f)$, making the DSS 15 thermal noise equally important as that of DSS 12. Thirdly, the contribution from the spacecraft oscillator turned out to be relatively small, because the oscillator phase-noise density decreases sharply with frequency, as $1/f^3$. Its effect on the phase error depends on $1 - H_1(f)$, which is a high-pass filter, attenuating all but an insignificant portion of the oscillator spectrum. This high-pass filter also limits the contribution of the DSS 12 VCO noise, which also varies as c/f^3. The contribution of the DSS 15 VCO phase noise was estimated to be small due to the small frequency multiplier following the VCO.

To estimate the symbol SNR degradation due to the root-mean-squared phase errors in the receiver, the statistical average of $\cos^2 \phi$ was used:

$$\text{Symbol SNR degradation} = < \cos^2 \phi > \qquad (3)$$

where ϕ represents the phase errors discussed above, with zero mean and a standard deviation equal to the total root-mean-squared error given in Table 2. By using this formula, the symbol SNR degradations can be estimated with and without carrier combining, as given in Table 2.

As for the DSS 15 receiver, similar calculations were made based on Eq. (A-25) of Appendix A. The results are presented in Table 3. Estimated degradations are also given.

C. Predicted Phase Errors in the 3-Way Mode

In the 3-way mode, the estimates given in Table 2 remain the same except for the spacecraft phase-noise contribution. In this case, the spacecraft phase noise was dominated by the effect of the transponder thermal noise in the uplink. The bandwidth of the phase-noise power spectrum was equal to that of the spacecraft carrier tracking loop, with one-sided bandwidth B_L of 54 Hz ($2B_{L0} = 21$ Hz at threshold), much larger than the ground receiver loop bandwidth. More precisely, the two-sided phase spectral density at the ground receiver input was $0.5\,K^2|H_{sc}(f)|^2 N_{osc}/P_{csc}$, where $H_{sc}(f)$ was the closed-loop transfer function of the spacecraft receiver, K was the downlink to uplink frequency ratio (240/221), N_{osc} was the spacecraft noise temperature (1039 K), and P_{csc} was the uplink carrier signal level received at the spacecraft (−135.5 dBm on May 3 and May 6, 1990). The spacecraft employed a predetection filter of 1900 Hz. By using these parameters in the third term of Eqs. (1) and (2), the phase error contributed by the uplink thermal noise is estimated to be 0.19 rad in either receiver, as noted at the bottom of Tables 2 and 3. This contribution caused an additional loss of −0.14 dB in both receivers.

D. Comparison Between Model and Observed Data

Table 4 provides a summary of the discussions above. The actual and expected performances under different conditions are compared. The discrepancies are within 0.2 dB in most conditions. Two inconsistencies are apparent in Table 4:

(1) The observed baseband-combining gain in the 3-way mode was significantly lower than the 1-way mode (1.4 versus 2.1 dB) or the prediction (2.3 dB). The phase noise coming down from the spacecraft, as caused by the uplink thermal noise, may have affected baseband-combining efficiency in a way not yet understood. Traditional calculation of baseband-combining optimal gain assumes the signals from different antennas to be correlated and noises to be uncorrelated. In this case, that assumption may not be valid because the noises received by these antennas from the same spacecraft were not necessarily uncorrelated. The more pronounced impact of correlated noise in a 3-way mode was likely caused by a wider noise spectrum of the spacecraft-tracking loop, as compared with the narrow $1/f^3$ spectrum in the 1-way mode. The effect of correlated noise from the spacecraft or the propagation media on baseband combining needs further investigation.

(2) Measurements conducted in the laboratory at JPL prior to the Goldstone demonstration showed a larger carrier-combining gain than had been predicted, as seen in Table 4. In the laboratory, two independent receiver channels with independent thermal noise sources and fiber-optic links were used to simulate conditions at DSS 12 and DSS 15. The baseband signal, from a single receiver or a resistive baseband combiner, was sent to the Compatibility Test Area (CTA21) for detection. The laboratory tests differed from the field demonstration in two respects: (a) Doppler was not simulated, the signal frequency was fixed throughout the test, and (b) no phase noise was introduced in the input carrier signal, only white Gaussian thermal noise was added. These two effects, however, did not explain the

0.9-dB difference in Table 4. This test was not repeated because the carrier-arraying modules had been removed from the receiver, which was needed for other purposes. It is suggested that the next carrier array experiment should examine the difference, if any, between laboratory test results and the analytical prediction of a model such as the one presented here.

E. Cycle Slips Observed in Laboratory Testing

Another indicator of carrier tracking loop performance near threshold is the frequency of cycle slips. This parameter was not measured at Goldstone because of time limitation, but the effect was simulated in the laboratory at 4-dB carrier margin. Carrier arraying significantly reduced the frequency of cycle slips, as shown in Fig. 6.

V. Conclusions

The demonstration conducted at Goldstone in May 1990 showed that an antenna with an inadequate G/T can indeed track a weak signal by carrier arraying with another antenna and produce telemetry baseband. Baseband-combining techniques can then be used to achieve a higher telemetry signal-to-noise ratio.

This demonstration marked the first time that carrier arraying was tried in the DSN to receive a spacecraft signal. The full benefit of combined aperture of two antennas was realized by using carrier arraying and baseband combining together. Lessons learned in this demonstration are applicable to the design of the next generation of DSN receivers.

Optimal design of a carrier-arraying system should consider all sources of phase errors, including both thermal and phase noises. An analytical model is presented to allow such evaluation. Predictions based on this model agree well with the Pioneer 11 data received, except for baseband-combining gain in the 3-way mode. Further investigation is needed to characterize the effect of correlated phase noise on the efficiency of baseband combining.

Acknowledgments

The authors are indebted to Robertson Stevens of the Telecommunications and Data Acquisition Office, whose encouragement made this demonstration possible. They also thank Clarence Hoynes and Joseph Johnson for operational support at Goldstone, and Peter Kinman and Stan Butman for many helpful discussions.

References

[1] M. H. Brockman, "Radio Frequency Carrier Arraying for High Rate Telemetry Reception," *DSN Progress Report 42-45*, vol. March and April 1978, Jet Propulsion Laboratory, Pasadena, California, pp. 209–223, June 15, 1978.

[2] D. Divsalar and J. H. Yuen, "Improved Carrier Tracking Performance with Coupled Phase-Locked Loops," *TDA Progress Report 42-66*, vol. September and October 1981, Jet Propulsion Laboratory, Pasadena, California, pp. 148–171, December 15, 1981.

[3] M. H. Brockman, "Performance Characteristics for an Array of Two Receiving Systems with Equal Apertures and Enhanced Radio Frequency Carrier Margin Improvement," *TDA Progress Report 42-84*, vol. October–December 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 112–126, February 15, 1986.

[4] R. Winkelstein, "Analysis of the Signal Combiner for Multiple Antenna Arraying," *Deep Space Network Progress Report 42-26*, Jet Propulsion Laboratory, Pasadena, California, pp. 102–118, January–February 1975.

[5] J. H. Yuen, ed., *Deep Space Telecommunications System Engineering*, JPL Publication 82-76, Jet Propulsion Laboratory, Pasadena, California, pp. 206–211, July 1982.

[6] V. A. Vilnrotter, W. J. Hurd, and D. H. Brown, "Optimized Tracking of RF Carriers With Phase Noise, Including Pioneer 10 Results," *TDA Progress Report 42-91*, vol. July–September 1987, Jet Propulsion Laboratory, Pasadena, California, pp. 141–157, November 15, 1987.

[7] R. Bunce, "Effect of VCO Noise on Phase-Lock Receiver," *JPL Space Programs Summary 37-61*, vol. 2, Jet Propulsion Laboratory, Pasadena, California, pp. 115–120, 1969.

Table 1. Pioneer 11 signal and noise conditions at Goldstone, 10-deg elevation, May 1990

Item	70-m, DSS 14	34-m STD, DSS 12	34-m HEF, DSS 15
P_c, dBm[a]	−167	−174	−174
T_{op}, K	30.5	33.5	47.5
N_0, dBm/Hz	−183.8	−183.3	−181.8
$2B_{L0}$, Hz	3	12[b]	12
Carrier margin[b] $(P_c/2N_0B_{L0})$, dB	12.0	−1.5[c]	−3.0

[a] Pioneer 10 signal level is about 5.5 dB lower than that of Pioneer 11.

[b] During the carrier arraying test, a 3-Hz bandwidth loop was used for DSS 12, with a carrier margin of 4.5 dB.

[c] Carrier margin for reliable signal acquisition and tracking should exceed 7 dB.

Table 2. Predicted DSS 12 receiver rms phase error, σ, contributed by various sources in the 1-way mode, rad

Source	Carrier not combined	Carrier combined
Thermal noise, DSS 12	0.53	0.29
Thermal noise, DSS 15	0	0.30
Phase noise, spacecraft	0.055	0.053
Phase noise, DSS 12 VCO	0.10	0.099
Phase noise, DSS 15 VCO	0	0.033
Total, rss	0.54	0.43
Symbol SNR degradation	−1.21 dB	−0.79 dB

Note: Spacecraft phase-noise effect would be 0.19 rad rms for the 3-way mode.

Table 3. Predicted DSS 15 receiver rms phase error, σ, contributed by various sources in the 1-way mode, rad

Source	Carrier not combined	Carrier combined
Thermal noise, DSS 12	0.51	0.52
Thermal noise, DSS 15	0.18	0.37
Phase noise, spacecraft	0.052	0.049
Phase noise, DSS 12 VCO	0.097	0.091
Phase noise, DSS 15 VCO	0.036	0.032
Total, rss	0.55	0.47
Symbol SNR degradation	−1.28 dB	−0.94 dB

Note: Spacecraft phase-noise effect would be 0.19 rad rms for the 3-way mode.

Table 4. Comparison between observed and predicted performance

Item	Observed	Predicted
Carrier-combining gain, dB		
1-way	0.31 ± 0.4	0.42
3-way	0.35 ± 0.5	0.40
Baseband-combining gain, dB		
1-way	2.15 ± 0.5	2.35
3-way	1.40 ± 0.5	2.35
Degradation from 1-way to 3-way, dB, DSS 12 or combined	−0.26 ± 0.5	−0.14
Laboratory test without Doppler or phase noise		
Carrier-combining gain, dB	1.27 ± 0.7	0.35
Baseband-combining gain, dB	2.10 ± 0.6	2.35

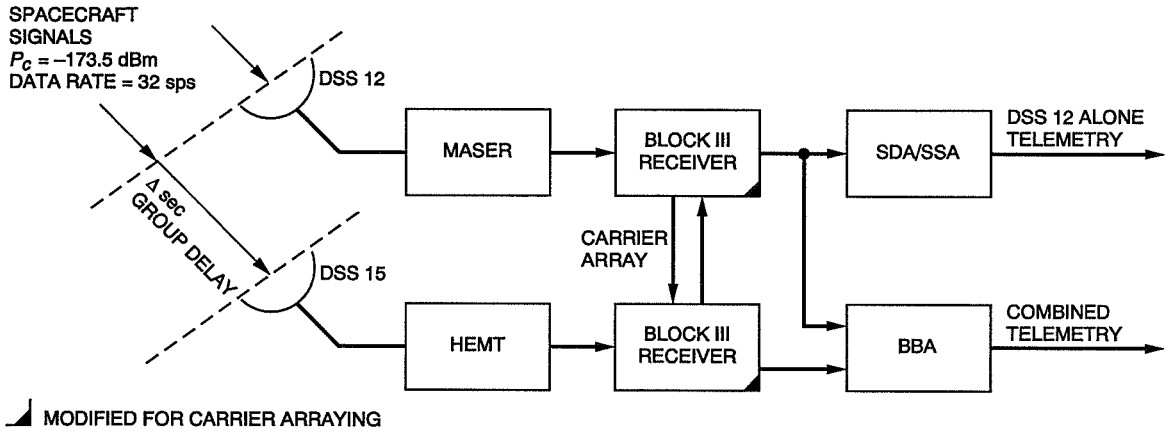


Fig. 1. Configuration of carrier-arraying demonstration.

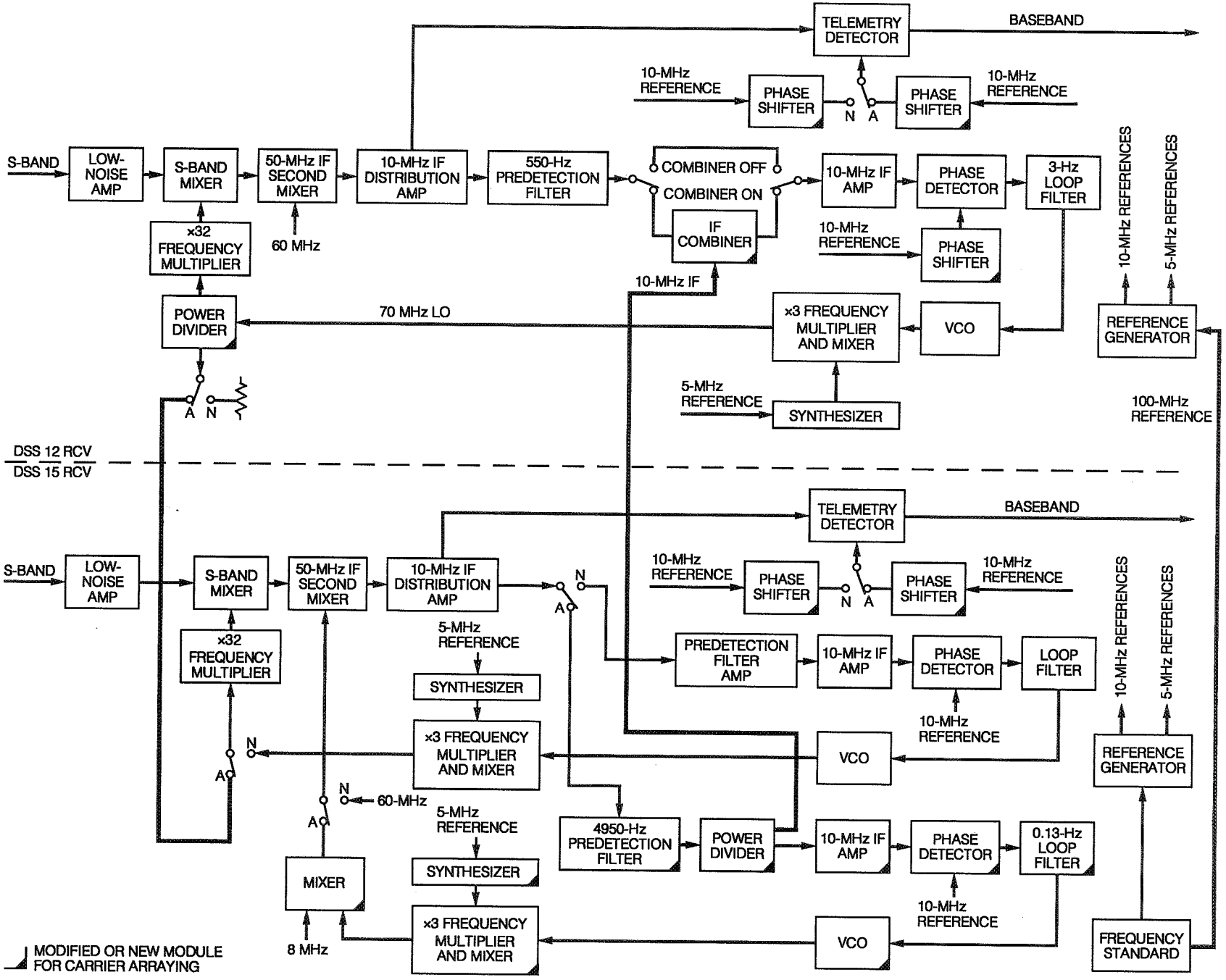


Fig. 2. Modifications of Block III receivers for carrier-arraying demonstration.

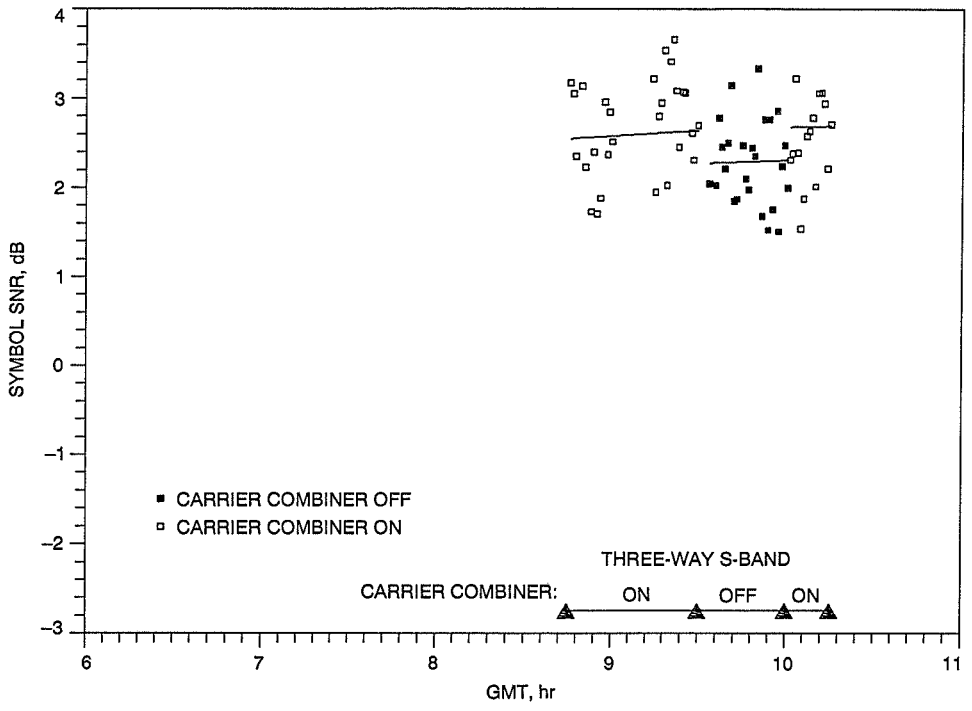


Fig. 3. Symbol SNR of combined telemetry on May 3, 1990.

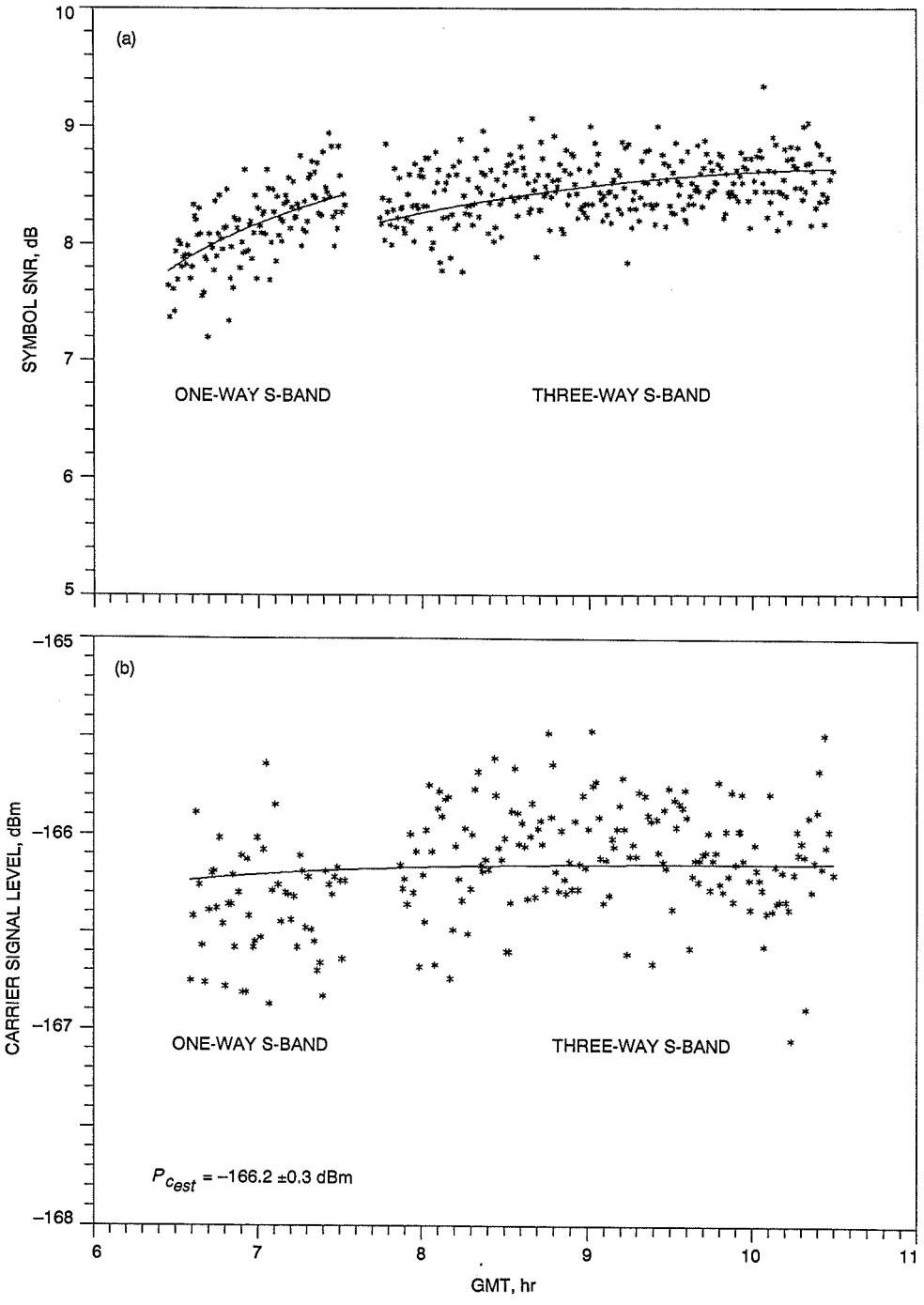


Fig. 4. May 3, 1990 DSS 14 measurements during Pioneer 11 track: (a) symbol SNR, and (b) carrier signal level.

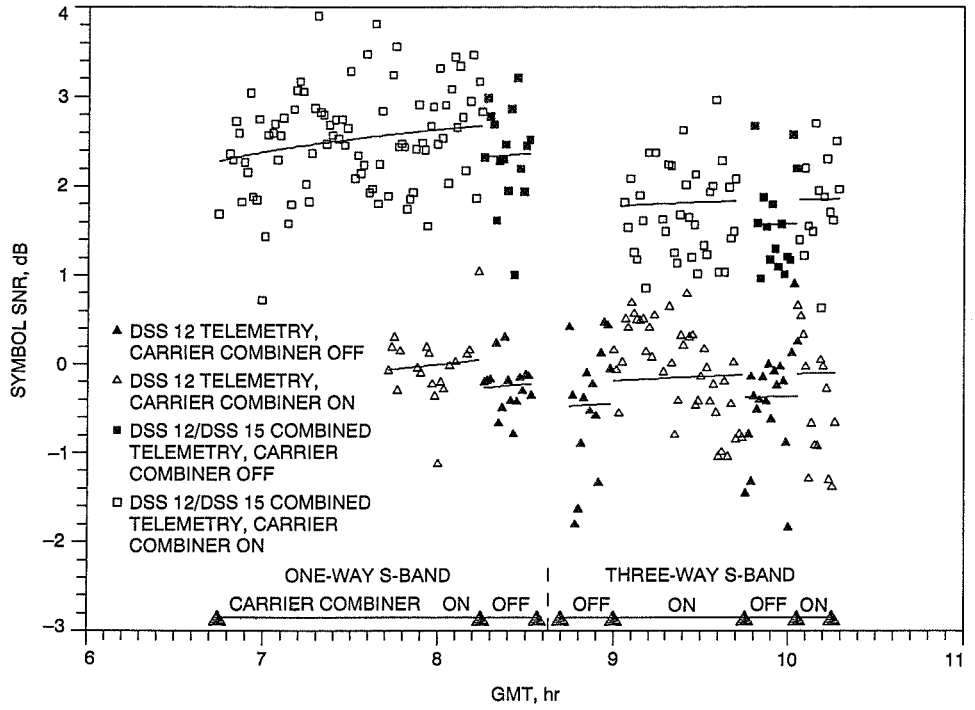


Fig. 5. Symbol SNR of single and combined telemetry on May 6, 1990.

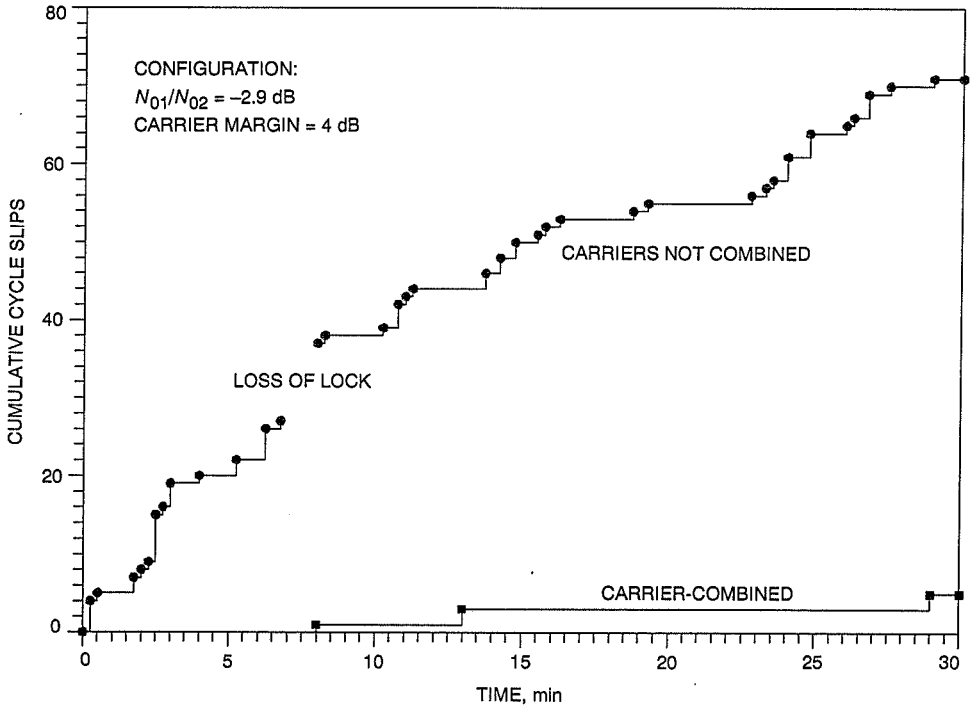


Fig. 6. Cycle slips measured in the laboratory on March 20, 1990.

Appendix A

The Effects of Spacecraft and Ground Receiver VCO Phase Noise on the Performance of Carrier-Arrayed Receivers

Consider the carrier-arraying system illustrated in Fig. A-1, its mathematical equivalent in Fig. A-2, and the equivalent with phase-noise effects in Fig. A-3. (These figures are taken directly from [2] where station and oscillator phase noise was not accounted for.) Assume now that each of the voltage-controlled oscillators (VCOs) has an associated phase noise, namely $\psi_{R_i}(t)$; $i = 1, 2, \ldots, N$. Then the instantaneous frequency $\dot{\theta}(t)$ of the ith VCO output, referenced to zero, is now related to the ith VCO input $z_i(t)$ by

$$\dot{\hat{\theta}}_i(t) = K_{VCO_i} z_i(t) + \dot{\psi}_{R_i}(t) \tag{A-1}$$

where a dot over a variable denotes differentiation with respect to time and K_{VCO_i} denotes the gain of the ith VCO. Letting p denote the Heaviside operator, that is, $px = dx/dt$, then in terms of the ith loop filter input $\tilde{S}_i(t)$, Eq. (A-1) can be rewritten as

$$\hat{\theta}_i(t) = \frac{K_{VCO_i}}{p}\left[F_i(p)\tilde{S}_i\right] + \psi_{R_i} \tag{A-2}$$

By expressing $\tilde{S}_i(t)$ in terms of the parameters characterizing the various loop input signals, then analogous to Eq. (A-1) in [2], the estimated carrier phase at the output of the VCO of station 1 is now

$$\hat{\theta}_1 = \frac{K_1 F_1(p)}{p}\left[\sqrt{P_1}\cos\theta_m \sin\left(\theta_1 + \psi_{T_1} - \hat{\theta}_1\right) + \sum_{i=2}^{N}\sqrt{P_i}\beta_i \cos\theta_m \sin\left(\theta_i + \psi_{T_i} - \hat{\theta}_i - \hat{\theta}_1\right) + \sum_{i=1}^{N}\beta_i N_i\right] + \psi_{R_1} \tag{A-3}$$

where

$\theta_i =$ the carrier phase at station i; $i = 1, 2, \ldots, N$

$\psi_{T_i} =$ the phase noise of the spacecraft signal received at station i; $i = 1, 2, \ldots, N$

$\beta_i =$ the weight associated with the signal from station i at the input to the carrier combiner; $\beta_1 = 1$

$\theta_m =$ the modulation index

$N_i =$ a Gaussian noise process with one-sided power spectral density N_{0i}; $i = 1, 2, \ldots, N$

$K_i =$ the total gain in the ith loop

Assume now that the time delay between the various stations is small as compared with the reciprocal of the bandwidth of the phase-noise process associated with the spacecraft signal. Then $\psi_{T_i} = \psi_T$ can be set for all $i = 1, 2, \ldots, N$. If the loop at station 1 is now linearized (i.e., assume that $\sin x = x$), then Eq. (A-3) simplifies to

$$\hat{\theta}_1 = \frac{K_1 F_1(p)}{p}\left[\left(\sqrt{P_1}\cos\theta_m\right)\left(\theta_1 + \psi_T - \hat{\theta}_1\right) + \sum_{i=2}^{N}\left(\sqrt{P_i}\beta_i\cos\theta_m\right)\left(\theta_i + \psi_T - \hat{\theta}_i - \hat{\theta}_1\right) + \sum_{i=1}^{N}\beta_i N_i\right] + \psi_{R_1} \tag{A-4}$$

Similarly, for station i; $i = 2, \ldots, N$ (after linearization),

$$\hat{\theta}_i = \frac{K_i F_i(p)}{p}\left[\left(\sqrt{P_i}\cos\theta_m\right)(\theta_i + \psi_T - \hat{\theta}_i - \hat{\theta}_1) + N_i\right] + \psi_{R_i} \tag{A-5}$$

Solving for $\hat{\theta}_i$ in Eq. (A-5) yields

$$\hat{\theta}_i = H_i(p)\left[(\theta_i + \psi_T - \hat{\theta}_1) + \frac{N_i}{\sqrt{P_i}\cos\theta_m}\right] + [1 - H_i(p)]\psi_{R_i}; \quad i = 2, 3, \ldots, N \tag{A-6}$$

where $H_i(p)$ is a closed-loop transfer function defined by

$$H_i(p) = \frac{K_i F_i(p)\sqrt{P_i}\cos\theta_m}{p + K_i F_i(p)\sqrt{P_i}\cos\theta_m}; \quad i = 1, 2, \ldots, N \tag{A-7}$$

Substituting Eq. (A-6) into Eq. (A-4) gives

$$\hat{\theta}_1 = H_1(p)\left[\theta_1 + \psi_T + \frac{N_1}{\sqrt{P_1}\cos\theta_m} + \sum_{i=2}^{N}\beta_i\gamma_i[1 - H_i(p)]\left(\theta_i + \psi_T - \hat{\theta}_1 + \frac{N_i}{\sqrt{P_i}\cos\theta_m} - \psi_{R_i}\right)\right] + [1 - H_1(p)]\psi_{R_1} \tag{A-8}$$

where $\gamma_i = \sqrt{P_i/P_1}$. Finally, letting $\Delta\psi_i = \psi_T - \psi_{R_i}$ represent the differential phase noise for the ith station and then solving Eq. (A-8) for $\hat{\theta}_1$,

$$\hat{\theta}_1 = \frac{H_1(p)\left[\theta_1 + \psi_T + \frac{N_1}{\sqrt{P_1}\cos\theta_m} + \sum_{i=2}^{N}\beta_i\gamma_i[1 - H_i(p)]\left(\theta_i + \frac{N_i}{\sqrt{P_i}\cos\theta_m} + \Delta\psi_i\right)\right]}{1 + \sum_{i=2}^{N}\beta_i\gamma_i H_1(p)[1 - H_i(p)]}$$

$$+ \frac{[1 - H_1(p)]\psi_{R_1}}{1 + \sum_{i=2}^{N}\beta_i\gamma_i H_1(p)[1 - H_i(p)]} \tag{A-9}$$

By comparing Eq. (A-9) with Eq. (A-7) in [2], Fig. A-2 can be immediately redrawn to include the effects of phase noise. The resulting equivalent linearized representation is illustrated in Fig. A-3.

Now the phase error in loop 1 is given by

$$\phi_1 = \theta_1 + \psi_T - \hat{\theta}_1 \tag{A-10}$$

Substituting Eq. (A-10) into Eq. (A-9) and simplifying gives the desired result:

$$\phi_1 = \frac{1 - H_1(p) + \sum_{i=2}^{N} \beta_i \gamma_i H_1(p)[1 - H_i(p)]}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(p)[1 - H_i(p)]} \theta_1 + \frac{1 - H_1(p)}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(p)[1 - H_i(p)]} \Delta\psi_1$$

$$- \frac{H_1(p)}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(p)[1 - H_i(p)]} \left[\frac{N_1}{\sqrt{P_1} \cos\theta_m} + \sum_{i=2}^{N} \beta_i \gamma_i [1 - H_i(p)] \left(\theta_i + \frac{N_i}{\sqrt{P_i} \cos\theta_m} - \psi_{R_i} \right) \right] \tag{A-11}$$

The mean-square phase jitter evaluation is provided below. The mean-square phase jitter in loop i; $i = 1, 2, \ldots, N$, is defined by

$$\sigma^2_{\phi_i} = E\Big\{ [\phi_i(s) - E\{\phi_i(s)\}] [\phi_i(-s) - E\{\phi_i(-s)\}] \Big\} \tag{A-12}$$

where E denotes statistical expectation and s is the Laplace transform operator. By using Eq. (A-11) in Eq. (A-12), one obtains for loop 1

$$\sigma^2_{\phi_1} = \frac{1}{2\pi j} \int \left| \frac{H_1(s)}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(s)[1 - H_i(s)]} \right|^2 ds \frac{N_{01}}{2P_1 \cos^2\theta_m}$$

$$+ \sum_{i=2}^{N} \beta_i^2 \gamma_i^2 \frac{1}{2\pi j} \int \left| \frac{H_1(s)[1 - H_i(s)]}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(s)[1 - H_i(s)]} \right|^2 ds \frac{N_{0i}}{2P_i \cos^2\theta_m}$$

$$+ \frac{1}{2\pi j} \int \left| \frac{1 - H_1(s)}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(s)[1 - H_i(s)]} \right|^2 S_{\Delta\psi_1}(s)\, ds + \sum_{i=2}^{N} \beta_i^2 \gamma_i^2 \frac{1}{2\pi j} \int \left| \frac{H_1(s)[1 - H_i(s)]}{1 + \sum_{i=2}^{N} \beta_i \gamma_i H_1(s)[1 - H_i(s)]} \right|^2 S_{\psi_{Ri}}(s)\, ds \tag{A-13}$$

where

$$\left.\begin{aligned} S_{\psi_{Ti}}(s) &= E\{\psi_{T_i}(s)\psi_{T_i}(-s)\} \\ S_{\psi_{Ri}}(s) &= E\{\psi_{R_i}(s)\psi_{R_i}(-s)\} \\ S_{\Delta\psi_i}(s) &= S_{\psi_{Ti}}(s) + S_{\psi_{Ri}}(s) \end{aligned}\right\} \tag{A-14}$$

are the two-sided power spectral densities associated with the phase-noise processes. In arriving at the third equation of Eq. (A-14), it has also been assumed that the transmit and receive phase-noise processes are uncorrelated. The first two terms of Eq. (A-13) are identical to Eq. (16) in [2]; the remaining two terms reflect the added degradation due to the various spacecraft and VCO phase-noise sources.

Under the assumption $H_i(s) = H_2(s)$; $i = 3, 4 \ldots, N$, Eq. (A-13) simplifies to

$$\sigma_{\phi_1}^2 = \frac{1}{2\pi j}\int \left|\frac{H_1(s)}{1 + H_1(s)[1 - H_2(s)](G-1)}\right|^2 ds \frac{N_{01}}{2P_1 \cos^2\theta_m}$$

$$+ \sum_{i=2}^{N} \beta_i^2 \gamma_i^2 \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G-1)}\right|^2 ds \frac{N_{0i}}{2P_i \cos^2\theta_m}$$

$$+ \frac{1}{2\pi j}\int \left|\frac{1 - H_1(s)}{1 + H_1(s)[1 - H_2(s)](G-1)}\right|^2 S_{\Delta\psi_1}(s) ds + \sum_{i=2}^{N} \beta_i^2 \gamma_i^2 \frac{1}{2\pi j}\int \left|\frac{H_1(s)[1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G-1)}\right|^2 S_{\psi_{Ri}}(s)\, ds \tag{A-15}$$

where

$$G = \sum_{i=1}^{N} \beta_i \gamma_i; \quad G - 1 = \sum_{i=2}^{N} \beta_i \gamma_i \tag{A-16}$$

The remaining step is to express Eq. (A-15) in terms of the loop dynamic parameters, such as damping and noise bandwidth. To do this, consider the case of second-order loops with imperfect integrating filters, i.e.,

$$F_i(s) = \frac{1 + \tau_{2i}(s)}{1 + \tau_{1i}(s)}; \quad \tau_{2i} << \tau_{1i} \tag{A-17}$$

Substituting Eq. (A-17) into Eq. (A-7) and recalling that the loop damping parameter r_i for the ith loop is defined by

$$r_i = \sqrt{P_i} \cos\theta_m K_i \tau_{2i}^2 / \tau_{1i} \tag{A-18}$$

one obtains

$$H_i(s) = \frac{r_i\left(\frac{1}{\tau_{2i}^2} + \frac{s}{\tau_{2i}}\right)}{s^2 + \frac{r_i}{\tau_{2i}}s + \frac{r_i}{\tau_{2i}^2}} = \frac{1 + \frac{r_i + 1}{4B_{Li}}s}{1 + \frac{r_i + 1}{4B_{Li}}s + \frac{1}{r_i}\left(\frac{r_i + 1}{4B_{Li}}\right)^2 s^2} \tag{A-19}$$

where B_{Li} denotes the loop noise bandwidth for the ith loop.

The phase error in loop i; $i = 2, 3, \ldots, N$, is given by

$$\phi_i = \theta_i + \psi_T - \hat{\theta}_1 - \hat{\theta}_i = \underbrace{\theta_1 + \psi_T - \hat{\theta}_1}_{\phi_1} + \underbrace{\theta_i - \theta_1 - \hat{\theta}_i}_{\phi_{ei}} \tag{A-20}$$

where ϕ_{ei} represents the phase error in loop i relative to that in loop 1. Using Eqs. (A-6)–(A-11), one obtains for loop i; $i = 2, 3, \ldots, N$,

$$\phi_i = (1 - H_i(p))\theta_i - D^{-1}H_1(p)(1 - H_i(p))\left(\theta_1 + \frac{N_1}{\sqrt{P_1}\cos\theta_m}\right)$$

$$- D^{-1}\left(H_i(p) + \beta_i\gamma_i H_1(p)(1 - H_i(p)) + H_i(p)\sum_{\substack{k=2\\k\neq i}}^{N}\beta_k\gamma_k H_1(p)(1 - H_i(p))\right)\left(\frac{N_i}{\sqrt{P_i}\cos\theta_m}\right)$$

$$- D^{-1}(1 - H_i(p))\sum_{\substack{k=2\\k\neq i}}^{N}\beta_k\gamma_k H_1(p)(1 - H_i(p))\left(\frac{N_k}{\sqrt{P_k}\cos\theta_m}\right) + D^{-1}(1 - H_i(p))(1 - H_1(p))\Delta\psi_1$$

$$- D^{-1}(1 - H_i(p))\left(1 + \sum_{\substack{k=2\\k\neq i}}^{N}\beta_k\gamma_k H_1(p)(1 - H_i(p))\right)\psi_{R_i} + D^{-1}(1 - H_i(p))\sum_{\substack{k=2\\k\neq i}}^{N}\beta_k\gamma_k H_1(p)(1 - H_i(p))\psi_{R_i} \tag{A-21}$$

where

$$D = 1 + \sum_{k=2}^{N}\beta_k\gamma_k H_1(p)(1 - H_i(p)) \tag{A-22}$$

Assuming again that $H_i(s) = H_2(s)$; $i = 3, 4, \ldots, N$, then analogous to Eq. (A-15), the total mean-squared phase error in loop i is

$$\sigma^2_{\phi_i} = \frac{1}{2\pi j}\int\left|\frac{H_1(s)[1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 ds\frac{N_{01}}{2P_1\cos^2\theta_m}$$

$$+ \frac{1}{2\pi j}\int\left|1 - \frac{[1 - H_2(s)][1 + H_1(s)[1 - H_2(s)](G_i' - 1)]}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 ds\frac{N_{0i}}{2P_i\cos^2\theta_m}$$

$$+ \sum_{\substack{k=2\\k\neq i}}^{N}\beta_k^2\gamma_k^2\frac{1}{2\pi j}\int\left|\frac{H_1(s)[1 - H_2(s)]^2}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 ds\frac{N_{0k}}{2P_k\cos^2\theta_m} + \frac{1}{2\pi j}\int\left|\frac{[1 - H_1(s)][1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 S_{\Delta\psi_1}(s)\,ds$$

$$+ \frac{1}{2\pi j}\int\left|\frac{[1 - H_2(s)][1 + H_1(s)[1 - H_2(s)](G_i' - 1)]}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 S_{\psi_{R_i}}(s)\,ds$$

$$+ \sum_{\substack{k=2\\k\neq i}}^{N}\beta_k^2\gamma_k^2\frac{1}{2\pi j}\int\left|\frac{H_1(s)[1 - H_2(s)]^2}{1 + H_1(s)[1 - H_2(s)](G - 1)}\right|^2 S_{\psi_{Rk}}(s)\,ds \tag{A-23}$$

where G and $G - 1$ are as in Eq. (A-16), and in addition

$$G_i' - 1 = \sum_{\substack{k=2 \\ k \neq i}}^{N} \beta_k \gamma_k \tag{A-24}$$

For $N = 2$, Eq. (A-23) simplifies to

$$\sigma_{\phi_2}^2 = \frac{1}{2\pi j} \int \left| \frac{H_1(s)[1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G-1)} \right|^2 ds \frac{N_{01}}{2P_1 \cos^2 \theta_m}$$

$$+ \frac{1}{2\pi j} \int \left| \frac{H_1(s)[1 - H_2(s)](G-1) + H_2(s)}{1 + H_1(s)[1 - H_2(s)](G-1)} \right|^2 ds \frac{N_{02}}{2P_2 \cos^2 \theta_m}$$

$$+ \frac{1}{2\pi j} \int \left| \frac{[1 - H_1(s)][1 + H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G-1)} \right|^2 S_{\Delta\psi_1}(s)\, ds + \frac{1}{2\pi j} \int \left| \frac{[1 - H_2(s)]}{1 + H_1(s)[1 - H_2(s)](G-1)} \right|^2 S_{\psi_{R2}}(s)\, ds \tag{A-25}$$

where now

$$G - 1 = \beta_2 \gamma_2 \tag{A-26}$$

Note that in the presence of the bandpass limiter in the receiver system, the limiter performance factor Γ, defined in Eq. (B-5) of Appendix B, should be included in the thermal contributions (first and second term) in Eqs. (A-15) and (A-25).

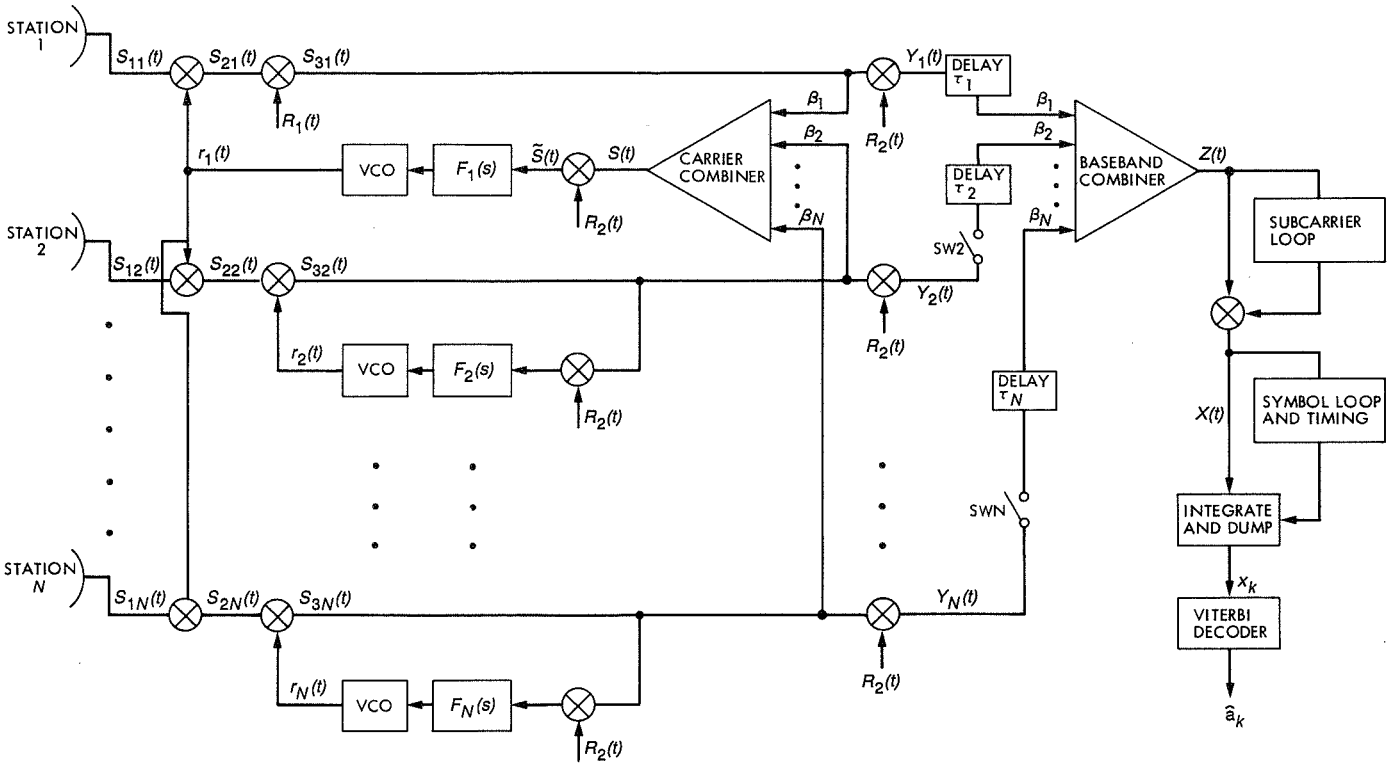


Fig. A-1. Configuration for arrayed network with carrier and baseband arraying.

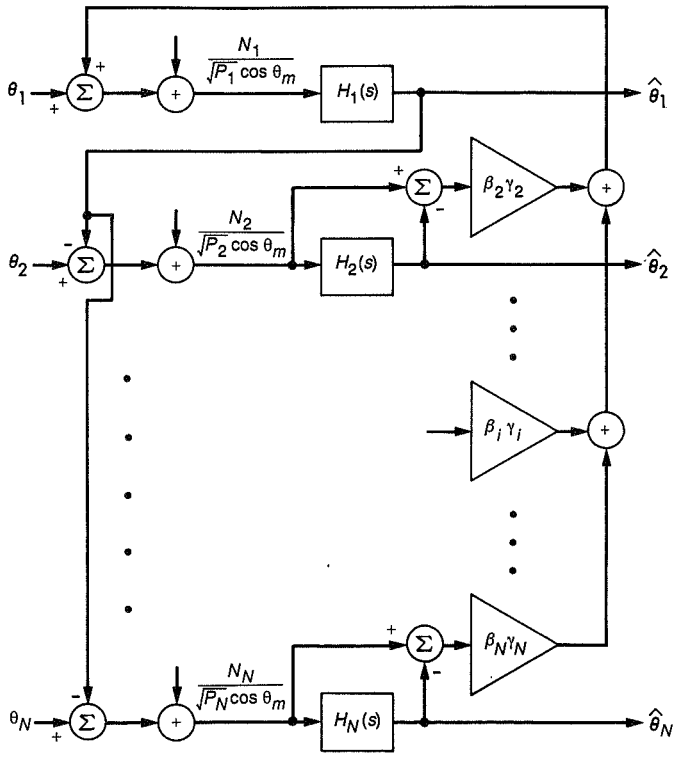


Fig. A-2. An equivalent linear representation of carrier-arraying system for phase jitter analysis.

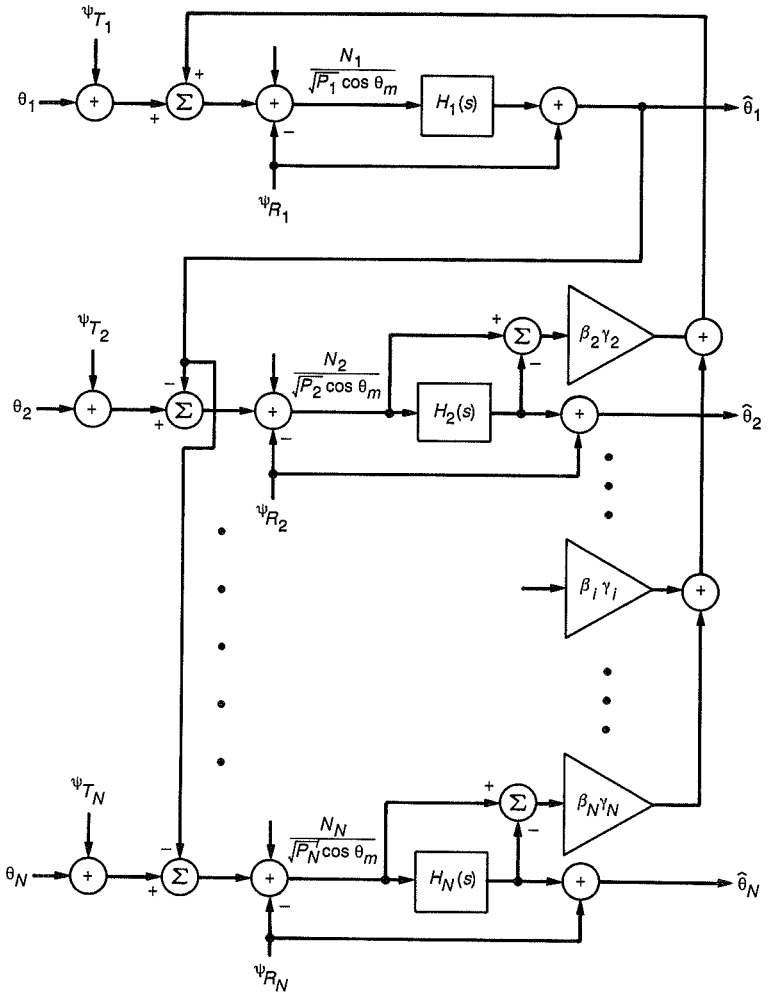


Fig. A-3. An equivalent linear representation of carrier-arraying system for phase jitter analysis (phase-noise effects included).

Appendix B

Optimum Design of Carrier-Arrayed Receivers With Different IF Filter Bandwidths

I. Background

A brief summary of the key results from [2], which is referenced later, is given here.

For $i = 1, 2, \ldots, N$, let β_i denote the weights applied to each signal entering the carrier combiner, P_{ci} denote the received carrier powers of the station input signals, B_{Li} denote the single-sided loop noise bandwidth of each receiver in the absence of arraying, and N_{0i} denote the noise power spectral densities of the receivers. Furthermore, for simplicity, assume that the loop gains and loop filters for stations $2, 3, \ldots, N$ are identical. This is tantamount to assuming that the receiver closed-loop transfer functions, $H_i(s)$, $i = 2, 3, \ldots, N$ are identical, or equivalently that $H_i(s) = H_2(s)$ for $i = 3, 4, \ldots, N$. Then, in the presence of carrier arraying, the mean-squared phase jitter of receiver 1 (the aided station) is given by

$$\sigma_{\phi 1}^2 = \frac{B_L \sum_{i=1}^{N} \beta_i^2 N_{0i}}{G^2 P_{c1}} \tag{B-1}$$

where

$$G = \sum_{i=1}^{N} \beta_i \gamma_i = 1 + \sum_{i=2}^{N} \beta_i \gamma_i \; ; \quad \gamma_i = \sqrt{\frac{P_{ci}}{P_{c1}}} \tag{B-2}$$

is the array "gain" and B_L is the loop bandwidth of receiver 1 in the presence of arraying, which is related to the loop bandwidth of receiver 1 in the absence of arraying B_{L1} by

$$B_L = \frac{1 + Gr_1}{1 + r_1} B_{L1} = \frac{1 + r}{1 + r_1} B_{L1} \tag{B-3}$$

Here, r_1 is the loop-damping parameter of receiver 1 in the absence of arraying, and $r = Gr_1$ is the same parameter in the presence of carrier arraying.

Consider now the effect of a bandpass limiter preceding each loop with IF bandwidths B_{IF_i}; $i = 1, 2, \ldots, N$. Equation (B-1) then becomes

$$\sigma_{\phi 1}^2 = \frac{\tilde{B}_L \Gamma \sum_{i=1}^{N} \beta_i^2 N_{0i}}{G^2 P_{c1}} \tag{B-4}$$

where Γ is the limiter performance factor approximately given by

$$\Gamma \cong \frac{1 + \rho_{in}}{0.862 + \rho_{in}} \tag{B-5}$$

and $\tilde{B}_L$ is the loop bandwidth with carrier arraying, which is given by

$$\tilde{B}_L = \frac{1 + \tilde{r}}{1 + r_1} B_{L1} \tag{B-6}$$

Here, ρ_{in} is the input SNR to the bandpass limiter of station 1, namely

$$\rho_{in} = \frac{G^2 P_{c1}}{\sum_{i=1}^{N} \beta_i^2 N_{0i} B_{IF_i}} \tag{B-7}$$

and $\tilde{r}$ is the loop damping of receiver 1 with carrier arraying, which is now given by

$$\tilde{r} = \frac{\sqrt{8/\pi^2}\tilde{\alpha}}{\sqrt{P_{c1}}} r_1 \tag{B-8}$$

with $\tilde{\alpha}$, the limiter suppression factor, approximated by

$$\tilde{\alpha} \cong \sqrt{\frac{0.7854\rho_{in} + 0.4768\rho_{in}^2}{1 + 1.024\rho_{in} + 0.4768\rho_{in}^2}} \tag{B-9}$$

Once the station receiver parameters have been defined for the unarrayed configuration, the mean-squared phase jitter of the arrayed system, as given in Eq. (B-4), can be minimized by proper selection of the combiner weights β_i; $i = 2, 3, \ldots, N$. The procedure for accomplishing this is discussed below. Note that the results described here differ from those discussed in [2] where it was assumed that

all IF filters had the same bandwidth. In the latter case it is possible to obtain an analytical characterization of the optimum weights, see Eq. (32) of [2]. Unfortunately, in the case of unequal IF filter bandwidths (e.g., the DSS 12 receiver has an IF bandwidth of 550 Hz whereas that of the DSS 15 receiver is 4950 Hz), it is not possible to analytically perform the optimization since the effective loop SNR is no longer a monotone increasing function of ρ_{in} as was assumed in [2].

II. Analysis

Without going through the details here, it was shown in [2] that when a bandpass limiter precedes the loop in each station (not indicated in Fig. 1), then the effective loop SNR, ρ_{L1}, of station 1 is given as the reciprocal of Eq. (B-4), namely,

$$\rho_{L1} = \frac{P_{c1}G^2}{\Gamma_1 \tilde{B}_{L1} \sum_{i=1}^{N} \beta_i^2 N_{0i}} \tag{B-10}$$

with Γ, $\tilde{B}_L$, ρ_{in}, and $\tilde{r}$ defined in Eqs. (B-5)–(B-8). These same parameters can be related to their values at threshold as follows. Letting the subscript "0" on a parameter denote its value at threshold, and defining the threshold condition (carrier arraying is assumed to be absent) by

$$\frac{P_{c10}}{2N_{01}\tilde{B}_{L0}} = 1 \tag{B-11}$$

then from Eq. (B-7), with $N = 1$ and $G = 1$,

$$\left.\begin{aligned} \rho_{in_{10}} &= \frac{P_{c10}}{N_{01}B_{IF_1}} = 2\frac{\tilde{B}_{L0}}{B_{IF_1}} \\ \tilde{\alpha}_0 &= \sqrt{\frac{0.785\rho_{in_{10}} + 0.4768\rho_{in_{10}}^2}{1 + 1.024\rho_{in_{10}} + 0.4768\rho_{in_{10}}^2}} \\ &\cong \sqrt{0.785\rho_{in_{10}}} \\ \tilde{B}_L &= \tilde{B}_{L0}\frac{1 + \tilde{r}_0\frac{\tilde{\alpha}}{\tilde{\alpha}_0}}{1 + \tilde{r}_0} \end{aligned}\right\} \tag{B-12}$$

Here, $\tilde{B}_{L0}$ and $\tilde{r}_0$ are, respectively, the values of loop bandwidth and damping factor associated with receiver 1 at threshold (no carrier arraying) and are specified in a given design, e.g., for DSS 12, $\tilde{B}_{L0} = 1.5$ Hz and $\tilde{r}_0 = 4$.

Since the radio loss associated with the data detection process is minimized by maximizing the loop SNR of the aided station, Eq. (B-10), one should select the set of weights $\{\beta_i;\ i = 2, 3, \ldots, N\}$ to achieve this goal. If the IF bandwidths of the receivers were all equal (as assumed in [2]), i.e., $B_{IF_i} = B_{IF}$ for all i, then one can find an analytical expression for these weights. The technique for doing this depends on recognizing that, for this case, ρ_{L1} is a monotonic increasing function of ρ_{in_1}; thus, maximizing ρ_{L1} with respect to the set of β_i's is equivalent to maximizing ρ_{in_1} with respect to the same parameters. When this is done, the following optimum set of weights results:

$$\beta_i = \gamma_i \frac{N_{01}}{N_{0i}}; \quad i = 2, 3, \ldots, N \tag{B-13}$$

While in principle, this same optimization can be accomplished for the case where the IF filter bandwidths are different, finding an analytical expression for the optimum weights is difficult. One reason is that ρ_{L1} is no longer a monotonic increasing function of ρ_{in_1} and thus one must directly maximize ρ_{L1} with respect to the set of β_i's.

To slightly simplify matters, consider the DSS 12/DSS 15 case where only two stations are involved, i.e., $N = 2$. Under these circumstances, Eq. (B-10) becomes

$$\rho_{L1} = \frac{P_{c1}}{\Gamma_1 N_{01}} \left[\frac{(1+\beta_2\gamma_2)^2}{1+\beta_2^2 K_{N2}}\right] \left(\frac{1+\tilde{r}_0}{\tilde{B}_{L0}}\right) \times \left(1 + \tilde{r}_0 \sqrt{\left(\frac{P_{c1}}{2N_{01}\tilde{B}_{L0}}\right)\left[\frac{(1+\beta_2\gamma_2)^2}{1+\beta_2^2 K_2 K_{N2}}\right]}\right)^{-1} \tag{B-14}$$

where

$$K_{N2} = \frac{N_{02}}{N_{01}}; \quad K_2 = \frac{B_{IF_2}}{B_{IF_1}} \tag{B-15}$$

Differentiating Eq. (B-14) with respect to β_2, keeping in mind that Γ is also a function of β_2 through ρ_{in_1}, see Eqs. (B-5) and (B-7), and equating the result to zero gives the optimum value of β_2. Finding the analytical solution for this optimum β_2 is difficult. One can simplify matters somewhat by noting that for small ρ_{in_1} (as is typically the case of interest), Γ_1 can be replaced by its value at

$\rho_{in_1} = 0$, namely $\Gamma_{10} = 1/0.862 = 1.16$. Secondly, when γ_2 is not too small, the term involving the square root in Eq. (B-14) will be reasonably large as compared with unity. Making these approximations, Eq. (B-14) simplifies to

$$\rho_{L1} = \frac{P_{c1}}{\Gamma_0 N_{01}} \left(\frac{1+\tilde{r}_0}{\tilde{B}_{L0}}\right) \left(\tilde{r}_0 \sqrt{\left(\frac{P_{c1}}{2N_{01}\tilde{B}_{L0}}\right)}\right)^{-1}$$

$$\times \left[\frac{(1+\beta_2\gamma_2)}{1+\beta_2^2 K_{N2}} \sqrt{1+\beta_2^2 K_2 K_{N2}}\right]$$

$$= C\left[\frac{(1+\beta_2\gamma_2)}{1+\beta_2^2 K_{N2}} \sqrt{1+\beta_2^2 K_2 K_{N2}}\right] \qquad \text{(B-16)}$$

where C is a constant that is independent of β_2. Thus, the optimum value of β_2 can be approximately obtained by maximizing the quantity

$$I = \left[\frac{(1+\beta_2\gamma_2)^2(1+\beta_2^2 K_2 K_{N2})}{(1+\beta_2^2 K_{N2})^2}\right] \qquad \text{(B-17)}$$

with respect to β_2. Before giving this result, note that if $K_2 = 1$ (i.e., equal IF bandwidths), then Eq. (B-17) reduces to

$$I = \left[\frac{(1+\beta_2\gamma_2)^2}{(1+\beta_2^2 K_{N2})}\right] \qquad \text{(B-18)}$$

which, when maximized with respect to β_2, yields the solution in Eq. (B-13), i.e., $\beta_2 = \gamma_2/K_{N2}$.

Returning now to Eq. (B-17) with $K_2 \neq 1$, then differentiating with respect to β_2 and setting the result equal to zero gives the solution

$$\beta_2^3 K_2 K_{N2}^2 - \beta_2^2 \gamma_2 K_{N2}(2K_2 - 1)$$

$$- \beta_2 K_{N2}(K_2 - 1) - \gamma_2 = 0 \qquad \text{(B-19)}$$

For the DSS 12/DSS 15 case, $B_{IF_1} = 550$ Hz and $B_{IF_2} = 4950$ Hz. Thus, $K_2 = 9$ and Eq. (B-19) becomes

$$9\beta_2^3 K_{N2}^2 - 17\beta_2^2 \gamma_2 K_{N2} - 8\beta_2 K_{N2} - \gamma_2 = 0 \qquad \text{(B-20)}$$

The table below evaluates the optimum β_2, as determined from Eq. (B-20), for various values of $\gamma_2^2 = P_{c2}/P_{c1}$ and $K_{N2} = 1$.

γ_2^2	β_2
1	2.256
0.5	1.793
0.1	1.245

Note that these optimum values of β_2 exceed unity, which was the value assumed in [2].

To determine the sensitivity of the receiver performance to the optimum choice of β_2, examine a plot of ρ_{L1} versus β_2 for a fixed set of system parameters. These parameters are usually determined for a given carrier margin M, which is defined as the operating signal-to-noise ratio in the threshold loop bandwidth with no arraying, namely,

$$M = \frac{P_{c1}}{2N_{01}B_{L0}} \qquad \text{(B-21)}$$

which, in view of Eq. (B-11), becomes

$$M = \frac{P_{c1}}{P_{c10}} \qquad \text{(B-22)}$$

i.e., the ratio of the operating carrier power to its threshold value. Using Eq. (B-12) in Eq. (B-14), the loop SNR of station 1, ρ_{L1}, can be written in terms of M as

$$\rho_{L1} = \frac{2M}{\Gamma_1}\left[\frac{(1+\beta_2\gamma_2)^2}{1+\beta_2^2 K_{N2}}\right](1+\tilde{r}_0)$$

$$\times \left(1+\tilde{r}_0\sqrt{M\left[\frac{(1+\beta_2\gamma_2)^2}{1+\beta_2^2 K_2 K_{N2}}\right]}\right)^{-1} \qquad \text{(B-23)}$$

where the value of ρ_{in_1} needed in Eq. (B-5) to compute Γ_1 is given in terms of M by

$$\rho_{in_1} = \left(M\frac{2\tilde{B}_{L0}}{B_{IF_1}}\right)\left[\frac{(1+\beta_2\gamma_2)^2}{1+\beta_2^2 K_2 K_{N2}}\right] \qquad \text{(B-24)}$$

Figure B-1 is a plot of ρ_{L1} versus β_2 for three values of carrier margin and the following parameters:

$\tilde{B}_{L0} = 1.5$ Hz

$r_0 = 4$

$B_{IF_1} = 550$ Hz and $B_{IF_2} = 4950$ Hz

$k =$ Boltzmann constant (1.38×10^{-23} W/Hz-K)

$T_1 = 25.6$ K and $T_2 = 25.6$ K

$$K_{N2} = \frac{N_{02}}{N_{01}} = \frac{kT_2}{kT_1} = 1$$

$$K_2 = \frac{B_{IF_2}}{B_{IF_1}} = 9$$

$$\gamma_2^2 = \frac{P_{c2}}{P_{c1}} = 1$$

Figure B-2 is a similar plot for a set of parameters typical of the current DSS 12/DSS 15 design at a 30-deg elevation angle:

$\tilde{B}_{L0} = 1.5$ Hz

$\tilde{r}_0 = 4$

$B_{IF_1} = 550$ Hz and $B_{IF_2} = 4950$ Hz

$k =$ Boltzmann constant (1.38×10^{-23} W/Hz-K)

$T_1 = 30.6$ K and $T_2 = 41.5$ K

$$K_{N2} = \frac{N_{02}}{N_{01}} = \frac{kT_2}{kT_1} = 1.356$$

$$K_2 = \frac{B_{IF_2}}{B_{IF_1}} = 9$$

$$\gamma_2^2 = \frac{P_{c2}}{P_{c1}} = 1$$

Tabulated below are the optimum values of β_2 and the corresponding maximum values of ρ_{L1} as determined from Figs. B-1 and B-2, respectively. Also included for comparison are the values of ρ_{L1} corresponding to $\beta_2 = 1$.

From Fig. B-1:

M	β_2	ρ_{L1}
7.7 dB	1.0	14.36
	1.9	15.92
5.7 dB	1.0	10.99
	1.85	12.10
3.7 dB	1.0	8.35
	1.8	9.13

From Fig. B-2:

M	β_2	ρ_{L1}
7.0 dB	1.0	12.35
	1.45	12.79
5.0 dB	1.0	9.39
	1.4	9.68
3.0 dB	1.0	7.08
	1.35	7.26

Thus, for the above two examples, it is possible to gain as much as a 0.5-dB improvement in loop SNR by an optimum choice of β_2.

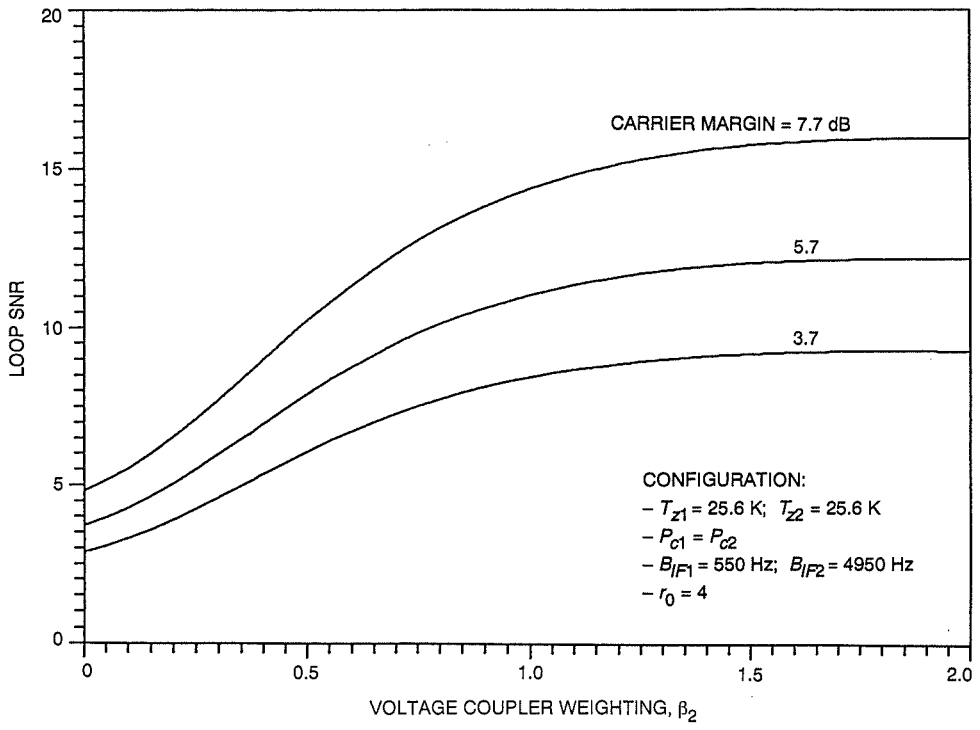


Fig. B-1. Loop SNR ρ_{L1} as a function of β and carrier margin.

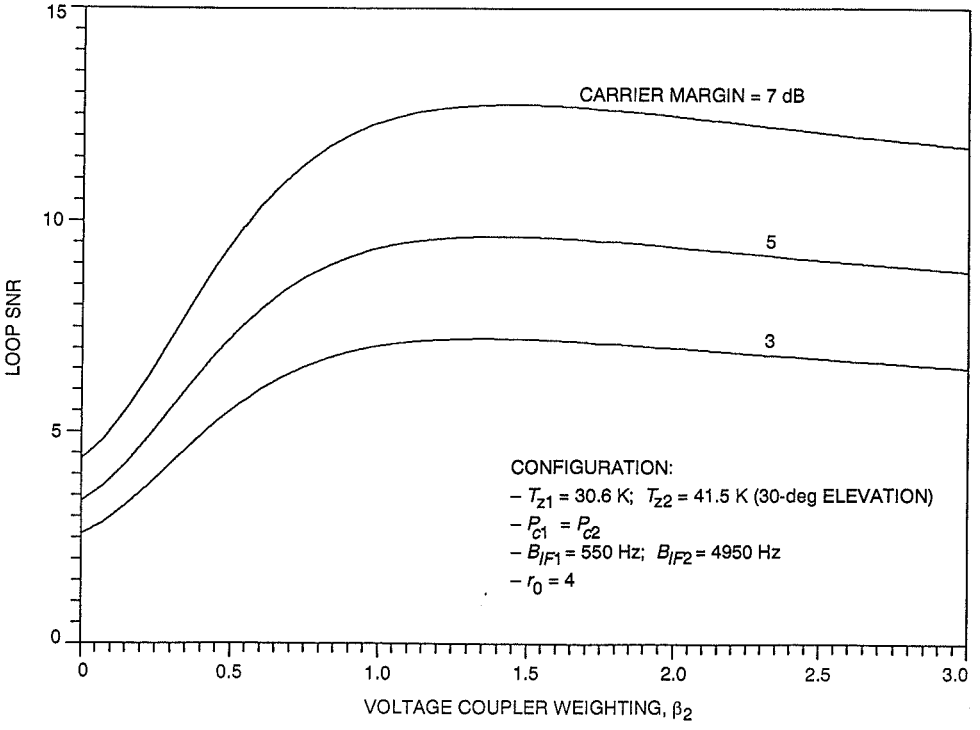


Fig. B-2. Loop SNR ρ_{L1} as a function of β and carrier margin, typical DSS 12/DSS 15 design at a 30-deg elevation angle.

ΔVLBI Data Performance in the Galileo Spacecraft Earth Flyby of December 1990

D. L. Gray
Navigation Systems Section

The performance of the delta very long baseline interferometry (ΔVLBI) data type during the recent Earth flyby of the Galileo spacecraft is presented. First, the data are given significance by showing why the mission design requires an accurate flyby. Then, the data's performance is analyzed, including its operational features, as well as its accuracy. Finally, the effectiveness of these data in improving the accuracy of the flyby is shown. Comparisons are made between the expected and actual performances of the data. The processing time, reliability, and accuracy of these data were very good, and a solid gain was made in the encounter accuracy.

I. Introduction

Delta very long baseline interferometry (ΔVLBI) has been used as an operational data type on each of the recent deep space projects, e.g., Voyager, Magellan, Ulysses, and Galileo. (An alternative name that is often used is ΔDOR, which stands for delta differential one-way range.) Recently, a number of hardware and software improvements have been made to improve the reliability, accuracy, and turnaround processing time required for this data type. A short history of the VLBI system and its improvements is given in Appendix A. A diagram of the data flow is shown in Fig. 1. It would be useful to have a report on the operational performance of the ΔVLBI data since these improvements have been made available.

The various deep space projects differ widely in their need for the ΔVLBI data type. Fortunately, the December 1990 Galileo flyby of the Earth provides an example where the recent improvements were used and the ΔVLBI data promised to be extremely useful for the Earth flyby. A study of the expected benefit of including ΔVLBI data during this encounter was performed earlier,[1] and a study of the effect of using connected-element interferometry data was reported in this publication [1]. This was the first of two Galileo flybys of the Earth, labelled Earth gravity assist 1 (EGA 1), which occurred on December 8, 1990, at a 960-km altitude. It yielded an unusual opportunity to assess the value of Earth-based radio data on the approach to Earth (i.e., the radio data became an encounter-relative data type). A series of ΔVLBI's was performed on the approach, and the purpose of this article is to assess the performance and usefulness of those ΔVLBI's in supporting the encounter. Where appropriate, the figures reproduce the original Galileo Navigation Team presentation material.

ΔVLBI data complement the basic radio Doppler data nicely because they are purely angular measurements (as seen from the Earth), while the Doppler data are most

[1] D. W. Murrow, "S-Band ΔDOR Data for Earth One Flyby," JPL Interoffice Memorandum GLL-NAV-89-60 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 8, 1989.

accurate in the radial direction. Since this article concerns ΔVLBI performance, only the two angular dimensions will be reported. For large distances out on an Earth encounter trajectory, these angular dimensions correspond to the classical B-plane[2] coordinates. Keep in mind that the high-gain antenna was not opened, and that the ΔVLBI's were of the single-frequency-band type at 2.3-GHz (S-band) frequency, rather than the dual-frequency-band type at 8.4 GHz/2.3 GHz (X-band/S-band), for which the spacecraft was designed. The 2.3-GHz data type is of additional interest given the recent problem in unfurling Galileo's high-gain antenna, since the same type of ΔVLBI data may be needed for later portions of the mission (especially given the likelihood that there would be little or no optical navigation data).

II. Mission Design

Because of the radioisotope thermoelectric generator (plutonium loaded) used to generate power for the spacecraft, an extra margin of safety was required against any chance of impinging on the Earth's substantial atmosphere. A typical mission requirement at each event on the encounter trajectory would be to keep the probability of impacting the encounter body to ≤ 0.01 (or equivalently ≤ 1.0 percent). Even this number is often relaxed somewhat if there would still be time (if necessary) to design and perform an emergency corrective maneuver before the encounter. However, for the two Galileo–Earth encounters, a requirement that the probability of impact always be $\leq 10^{-6}$ was levied, and the consequences were analyzed in the *Galileo Navigation Plan*[3]. This very conservative requirement led to a sequence of three intermediate-biased aimpoints and six trajectory correction maneuvers (TCM's) on the approach trajectory. The resulting maneuver spacings and altitudes are shown in Fig. 2, and the sequence of B-plane aimpoints is shown in Fig. 3.

Since the Earth flyby was used to shape the spacecraft's trajectory, any spatial miss in the encounter would require a significant propellant penalty to correct later. Propellant penalty has been defined for Galileo as the increase in propellant required to complete the mission with 90-percent confidence. The propellant penalty for any miss was calculated, given the maneuvers planned after the Earth flyby, to produce contours of constant propellant penalty in the Earth B-plane. The constant propellant penalty contours are very flattened in the radial direction, due primarily to the high cost of correcting the resulting timing errors. As a rough estimate of the cost, in the radial direction a 10-km miss in the B-plane requires a 10 m/sec penalty in propellant to correct after the flyby, while in the tangential direction a 10-km miss only requires a 2 m/sec penalty to correct. Since the Galileo propellant budget is very tight, even a 10 m/sec fuel savings will increase the confidence level of completion of the nominal Jupiter tour or, alternately, increase the length of a possible extended mission.

A timeline of the ΔVLBI's scheduled and performed is given in Fig. 4. Note that the ΔVLBI data in support of a given maneuver stop well before the date of the maneuver, to allow time for design and implementation.

III. ΔVLBI Performance

A. Processing Time

There are several measures to be used in assessing ΔVLBI performance as an operational data type. The simplest of these is the processing delay time. Twelve hours were allocated, following the completion of acquisition, for delivery of processed data to the Project navigation computer. This total allocation was broken down as follows: 2.5 hr for playback, 5 hr for correlation, 1 hr for conditioning and delivery to the Navigation Team, and 3.5 hr for internal margin. Actual delay times were about 2 hr for the playback, 4 hr for the correlation, and 10 min for the conditioning and delivery—well within the requirement. Actually, processing time of about this length was demonstrated in 1982, but was not a requirement until the Magellan Project. The Voyager Project ΔVLBI data-processing time was typically several days, in order to save money.

B. Scheduling Requirements

A second measure of performance is the difficulty of scheduling a ΔVLBI observation, as judged by the requirements the scheduling places on the Deep Space Network. One factor is the amount of Deep Space Station time required. Table 1 shows the steps and time intervals required to perform a single ΔVLBI observation, which, even if no other data are to be taken, adds up to nearly two hours of dedicated time for each of two antennas. Note that, for most of this time, it is either impossible or impractical to be receiving data from the spacecraft. (By comparison, when Doppler and ranging data are taken, similar pre- and post-calibration times are required, but other

[2] The B-plane is perpendicular to the direction of the incoming spacecraft trajectory asymptote. Its origin is at the center of the encounter body, and the B-plane coordinates of the spacecraft encounter are computed on the trajectory asymptote (as though the encounter body did not cause any bending).

[3] *Galileo Navigation Plan*, Project Document 625-566, Revision A (internal document), Jet Propulsion Laboratory, Pasadena, California, October 1989.

downlink data are normally received while the navigation data are obtained.) Another requirement is to perform the observation during the relatively short period of station overlap, which makes it impossible to adjust the start time by any large amount when scheduling conflicts occur. To make matters worse, with the Galileo high-gain antenna unavailable, it has been necessary to use the large antennas of the 70-m subnet for these observations. Finally, the Galileo sequence command charging algorithm is such that nine commands are charged to the navigation total for each ΔVLBI carried out (or five commands for the later observations, which were performed without turning off the telemetry).

Since DSN antenna time is severely oversubscribed, it is necessary to justify that these data are needed in spite of the relatively difficult requirements for scheduling the time. Often the basis for the need is not readily apparent and must be defended in detail. An example of this is a request for a specific pair of observations by P. Kallemeyn and V. Pollmeier,[4] the text of which is included as Appendix B. Even after the schedule has been agreed to, there are likely to be unexpected cancellations. On the timeline of Fig. 4, each cancelled ΔVLBI observation is indicated. The six cancelled observations were lost to the high-priority needs of other projects, which means these observations did not have a high enough priority to survive, given the competition for deep-space antenna time. The scheduling is done in an interactive, iterative fashion, so to some extent early cancellations were replaced by using requests for later data. However, as discussed later, an important result of using ΔVLBI data may have been weakened by these cancellations. Fortunately, the desirability of some ΔVLBI observations was clear in this case, otherwise even more might have been cancelled.

C. Reliability

The reliability of the ΔVLBI data has an indirect effect on the ease of scheduling, since extra observations have to be scheduled to make up for any expected loss rate. Figure 4 also indicates the four or five[5] ΔVLBI observations that failed, out of 27 attempts. Probably the fairest assessment of reliability would be to assume four failures out of the 27 attempts, or 85-percent reliability. This 85-percent reliability is quite good, and it will be interesting to see whether future ΔVLBI's will maintain such a rate.

D. Accuracy

Achieved accuracy of the data is the best known measurement of performance. First, a detailed orbit estimate is made, estimating the spacecraft's initial state, its accelerations, and any other parameters that may affect the values of the observables (which are Doppler, range, and ΔVLBI measurements). When this is done, a test of the quality of the result is obtained by comparing the actual measurements taken with those measurements that would be predicted by using the fitted parameter values. The differences among these predicted and actual measurements are called residuals. These are plotted for the earlier ΔVLBI observations in Fig. 5, resulting in a standard deviation of 0.20 m.[6] Since Doppler data were also fitted in the computer run used to obtain Fig. 4, the ΔVLBI scatter is somewhat conservative as a measure of its precision, but does not include any bias error. In order to discuss the performance of these ΔVLBI observations, it would be useful at this point to have an analytic estimate of the expected accuracy.

Three analytic estimates of the expected ΔVLBI accuracy for Galileo have been made. The first is from the *Galileo Navigation Plan*, assuming the system as designed (with a high-gain antenna and dual frequency with 8.4 GHz), and it results in an expected scatter of 0.14 m with a total accuracy of 0.30 m. After it was realized that some observations would be made with the present spacecraft configuration, a second estimate was made in May 1990[7] for this EGA 1 approach under a variety of possible conditions. For the most likely conditions, with a conservative assumption of 0.60 m on the expected ionospheric error, the component uncertainties are shown in Table 2, and the result was an expected total error of 0.74 m. An ongoing general effort is made to improve the ΔVLBI data type (see e.g., [2]), and after this accuracy estimate was published, a specific effort was made to refine and im-

[4] P. Kallemeyn and V. Pollmeier, "Rationale for a ΔDOR Pair Two Days After TCM7," JPL Interoffice Memorandum GLL-NAV-90-051 (internal document), Jet Propulsion Laboratory, Pasadena, California, June 7, 1990.

[5] The last east–west baseline observation was delayed and about to fail when J. Border supplied, in real time, an alternate quasar with later viewing available. This observation was not used in the time-pressured orbit estimate that immediately followed (due to setup time and additional analysis needed to verify use of the new quasar), but was used in later reconstruction operations.

[6] A short explanation is needed here since ΔVLBI errors may be expressed in any of three related dimensions: nanoseconds, meters, or nanoradians, as diagrammed in Fig. 6. The measurement is the spacecraft delay (difference in signal arrival time between the two antennas) minus the quasar delay, so the most direct units are nsec. However, using an approximate conversion of 6×10^6 m for the projected length of the baseline and 0.3 m/nsec for the speed of light makes a typical accuracy of 1.0 nsec equivalent to 0.3 m or 50 nrad.

[7] J. S. Border, "ΔDOR Observation Parameters for 1990 Galileo Earth Approach," JPL Interoffice Memorandum 335.1-90-025 (internal document), Jet Propulsion Laboratory, Pasadena, California, May 10, 1990.

prove the calculation of the dominant ionospheric delay effects [3].[8,9] The new analysis predicted an ionospheric error of only 0.09 m. This change alone would reduce the expected total error to 0.45 m, but a better understanding of instrumental effects, plus the use of 70-m (rather than 34-m) antennas further reduced the expected error. After the encounter, a third analytic estimate of the expected errors was made,[10] with the component uncertainties also included in Table 2, for a total expected error of 0.21 m.

The third analytic estimate (of 0.21-m uncertainty) is very close to the observed scatter (of a 0.20-m standard deviation). Note that these both omit the bias uncertainty due to the quasar position error. Taking these factors into account, the observed scatter of 0.20 m and the *Galileo Navigation Plan* estimate of a 0.30-m total uncertainty will be used as best estimates of the achieved single-observation uncertainty. The resulting single-observation uncertainties (along the baseline) of the ΔVLBI data arc are shown in Table 3.

With a pair of observations (one N–S plus one E–W observation), one can estimate the spacecraft position in two dimensions. Since the two baselines are not orthogonal, the resulting two-dimensional accuracy is a function of the direction in the B-plane, as shown in Fig. 7.[11] The intent of Fig. 7 is to supply multipliers for use in extending the single-dimension scatter and uncertainty accuracies given above (or the corresponding encounter accuracies from Table 3) to two dimensions. A pair of observations will have an uncertainty ranging from 0.77 times the single-observation uncertainty (in the most favorable direction) to 1.85 times the single-observation uncertainty (in the direction of maximum uncertainty). For example, at the end of the usable data arc, Table 3 gives the accuracy of an observation as equivalent to 0.85 km mapped to encounter. A pair of these observations has a maximum uncertainty of 1.85 × 0.85 km, or 1.6 km. Unfortunately, the direction of maximum uncertainty corresponds approximately to the radial direction for the EGA 1 flyby aimpoint, which is bad for both propellant cost and safety. However, a radial spacecraft delivery error of 1.6 km would result in no noticeable increase in impact danger and a 1.0-m/sec propellant penalty. For comparison, a study from the *Galileo Navigation Plan* of the effectiveness of the Doppler data is shown in Fig. 8. For the same data cutoff time, the worst direction ($B \cdot R$) Doppler uncertainty is 12.9 km, more than eight times that of a ΔVLBI pair. Next, the effect on encounter accuracy must be discussed.

IV. Effect on Encounter

The a posteriori reconstructed delivery error of 8.4 km is plotted in Fig. 9, together with the computed (solid line) and adjusted (dashed line) delivery uncertainty ellipses. (Due to known difficulties in fitting the orbit data, an adjustment factor of 1.5 was used to enlarge the computed uncertainty ellipse.) An attempt was also made to construct what would have happened if no ΔVLBI data had been available, and the resulting Doppler-only error and uncertainty ellipses have also been plotted in Fig. 9. First, note that the size of the reconstructed delivery error is much larger than the ΔVLBI maximum final position uncertainty of 1.6 km, although its direction is near to the maximum ΔVLBI uncertainty direction shown in Fig. 7. Thus, there would appear to be a significant growth in error due to propagating the trajectory from the last ΔVLBI observations to the encounter. The size of the delivery error was also larger than the computed encounter uncertainty ellipse. This behavior suggests the possibility of an inadequately modelled error source, which is thought to be primarily the sum of small errors in modelling solar pressure effects plus Trajectory Correction Maneuver 7 (TCM7) modelling errors. A discussion of this topic is given below.

A. Estimating Acceleration and Velocity Changes

Both Doppler and ΔVLBI data can be affected by many possible error sources, but there is an important difference between the two. While ΔVLBI is a direct measure of angle, Doppler depends on its time history to infer position in the B-plane. Since a time history is also needed to estimate any accelerations or velocity changes acting on the spacecraft, it is difficult to estimate any velocity changes or accelerations in the B-plane by using Doppler data alone. In fact, an inadequate modelling of the solar pressure and other nongravitational accelerations led to a similar problem with the orbit determination delivery 27 (OD 27) estimate, just before the Earth encounter ΔVLBI data were initiated, as shown in Fig. 10.

[8] A. J. Mannucci, "Checking Faraday Ionosphere Calibration Using TEMPO Passes," JPL Interoffice Memorandum 335.1-90-025 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 16, 1990.

[9] A. J. Mannucci, "Temporal Statistics of the Ionosphere," JPL Interoffice Memorandum 335.1-90-056 (internal document), Jet Propulsion Laboratory, Pasadena, California, October 25, 1990.

[10] J. S. Border, "Galileo S-Band ΔDOR Measurement Accuracy for First Earth Approach, October–November 1990," JPL Interoffice Memorandum 335.1-91-012 (internal document), Jet Propulsion Laboratory, Pasadena, California, April 23, 1991.

[11] For an Earth encounter asymptote, the angles made by the baselines (as seen from the spacecraft) correspond to B-plane angles. These baseline angles are not constant, but only vary by a few degrees over the course of any observation and also over the period of the ΔVLBI data arc.

One of the things it was hoped the ΔVLBI data would do was to better estimate these accelerations, so it is fair to ask why a better job was not done. An answer can be seen from an analysis of Fig. 4. The original planning for ΔVLBI scheduling provided a time interval of 70 days before the orbit data cutoff for TCM7 (with an increased amount of data in the last 23 days). This was the only opportunity to use ΔVLBI data to estimate the nongravitational accelerations with a minimum of other perturbations (such as TCM's, science turns, and momentum corrections) in the data arc. Failure of the first two observations reduced the available time interval to 55 days, which is only a small loss. However, the cancellation of six sequential E–W baseline observations reduced the available time interval with pairs of observations to just 11 days. When the last N–S observation in this group failed, the effective time interval was reduced to just 8 days.

A sample calculation is shown in Appendix C to illustrate how drastically the short effective time interval reduces accuracy when estimating an acceleration. Using realistic times and ΔVLBI measurement accuracies leads to a resulting encounter uncertainty of 13.7 km due to the poorly estimated uncertainty—similar to the size of the actual encounter error. This is a rather simplistic example. In reality, the encounter error is thought to be a combination of several velocity and acceleration estimate errors. However the example does show how drastically the ability to estimate the solar pressure effect can be reduced by shortening the measurement arc.

The importance of the cancelled E–W ΔVLBI observations can also be seen by comparing the encounter error and covariance ellipse in Fig. 9 with the ΔVLBI uncertainty plot in Fig. 7. The encounter covariance ellipse is narrower than the ΔVLBI uncertainty plot, due to the shortage of E–W observations relative to N–S observations. Furthermore, the ellipse's semimajor axis (and also the encounter error itself) points much closer to the anti-N–S baseline direction than to the maximum ΔVLBI uncertainty direction. Thus, the trajectory was well resolved along the N–S baseline, but much less well resolved in the direction perpendicular to it. That this should be true in spite of an even distribution of N–S and E–W observations in the latter portion (where the observations are most accurate) attests to the predominance of velocity and acceleration estimation errors in this encounter.

B. Comparison With Doppler Solution

In June 1990, when this campaign of ΔVLBI data was recommended, it was calculated that the expected increase in propellant margin would be 10 m/sec (10 m/sec is equal to 9.4 kg of propellant). By comparison, based on the actual delivery, the increase in propellant margin reported by the Navigation Team was 5 m/sec. While smaller than the expected value, this increase is within the range of results that would be expected in drawing one sample from the distribution of likely encounter deliveries. To show how this encounter might have fared without the ΔVLBI data, the effect of using the best Doppler-only orbit solution was plotted in Fig. 9. Even though the Doppler-only delivery error was almost three times as large (at 22 km), it was in a direction where corrections would have a low propellant cost. Thus, the savings would have been only about 5 m/sec. However, if the ΔVLBI data had been about three times as effective in estimating the spacecraft orbit in a direction perpendicular to the N–S baseline, then a 10-m/sec propellant savings would have been realized.

A final insight into the orbit estimation process can be gained by analyzing the ΔVLBI residuals at the second (TCM8) orbit cutoff, as shown in Figs. 11 and 12. Figure 11 shows the result when only the Doppler data are used to estimate the orbit. In this case, the ΔVLBI residuals drift off rather badly, reaching values of about 15 m by the end of the data arc. This shows that the Doppler data are not very good for estimating the changes in velocity in the B-plane. Notice that the drifting becomes especially bad after TCM7. By contrast, after adding in the ΔVLBI observations to the fitting process, the residuals are much smaller, as shown in Fig. 12. However, these residuals are still noticeably larger than the set obtained for the earlier cutoff, as shown in Fig. 5, indicating some difficulty in fitting the additional changes in velocity now included in the data arc. It is apparent that the E–W baseline residuals show some remaining trend after the fitting process, but it was not obvious at the time how to eliminate this trend. Keep in mind that a very limited time period was available for the orbit determination processing so that the design of TCM8 could proceed on time. After the fact, it soon began to appear that a better solution could have been made by restricting any variation in the reconstruction of the TCM7 ΔV, and forcing an additional adjustment to be made in either the solar pressure terms or some other nongravitational acceleration term.

V. Conclusions

A mostly successful ΔVLBI campaign was carried out in support of the Galileo EGA 1 encounter. The processing time, reliability, and accuracy of these data were very good, and a solid gain was made in the encounter accuracy. The final result would probably have been even better if it were not for the cancellation of all the early observations on the E–W baseline.

Acknowledgments

Deep thanks are due to the Galileo Navigation Team, without whose wholehearted cooperation this article would have been impossible, and especially to Frank Nicholson and Peter Kallemeyn of that Team, who provided handout materials and discussions. Earl Maize and Chris Potts of the Galileo Navigation Team provided additional materials on fuel usage. Jim Border of the Tracking Systems and Applications Section provided discussions plus ΔVLBI expected performance estimates. Finally, Jim McDanell, Carl Christensen, and Frank Nicholson, all of the Navigation Systems Section, as well as Jim Border, provided helpful reviews of the article.

References

[1] S. W. Thurman, "Galileo Earth Approach Navigation Using Connected Element Interferometer Phase-Delay Tracking," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 34–47, February 15, 1990.

[2] J. S. Ulvestad, O. J. Sovers, and C. S. Jacobs, "A Higher Density VLBI Catalog for Navigating Magellan and Galileo," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 274–300, February 15, 1990.

[3] J. S. Border, W. M. Folkner, R. D. Kahn, and K. S. Zukor, *Precise Tracking of the Magellan and Pioneer Venus Orbiters by Same-Beam Interferometry*, paper AAS 91-191, AAS/AIAA Spaceflight Mechanics Meeting, Houston, Texas, February 11, 1991.

Table 1. DSS time requirement for ΔVLBI support

1. Load MDA and station-pointing predictions	45 min
2. DSP pre-calibration	30 min
3. Ranging pre-calibration	30 min[a]
4. ΔVLBI collection	
Spacecraft observation	5 min[b]
Slew	3 min[c]
Quasar observation	10 min[b]
Slew	3 min[c]
Spacecraft observation	5 min[b]
5. Post-calibration	15 min
Total	~2 hr (2.5 hr if with ranging)

[a] Only if Doppler and ranging data are to be taken. Not needed if this is a stand-alone pass.

[b] Minimum observation time is determined by accuracy requirement and signal power level.

[c] Required slew time depends on the separation between the spacecraft and quasar.

Table 2. Galileo S-band ΔDOR error budget for first Earth approach: original and revised (1σ)

Error Source	May 1990 Prediction, cm	Revised, cm
Spacecraft signal-to-noise ratio (SNR)	14	6
Quasar SNR	26	9
Phase ripple	31	15
Baseline	2	2
Troposphere[a]	9	4
Ionosphere[a]	60	9
Solar plasma[a]	1	1
Root sum square	74	21

[a] Sometimes combined and called media.

Table 3. ΔVLBI data characteristics

	Days to Encounter	Range to Earth, 10^6 km	Scatter,[a] km	Accuracy,[b] km
Data Start	90	87	2.90	4.35
	48	43	1.43	2.14
	40	35	1.17	1.76
Data End	20	17	0.57	0.85

[a] Scatter = 0.20 m × range/baseline, where baseline = 6×10^6 m.

[b] Accuracy = 0.30 m × range/baseline.

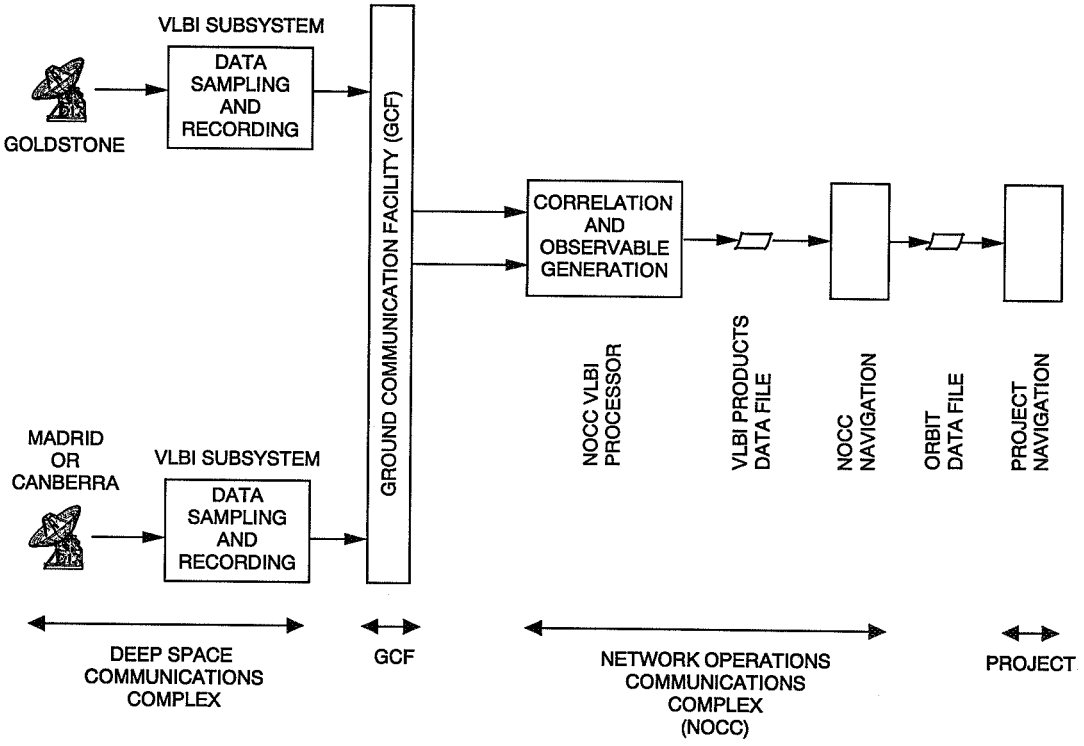
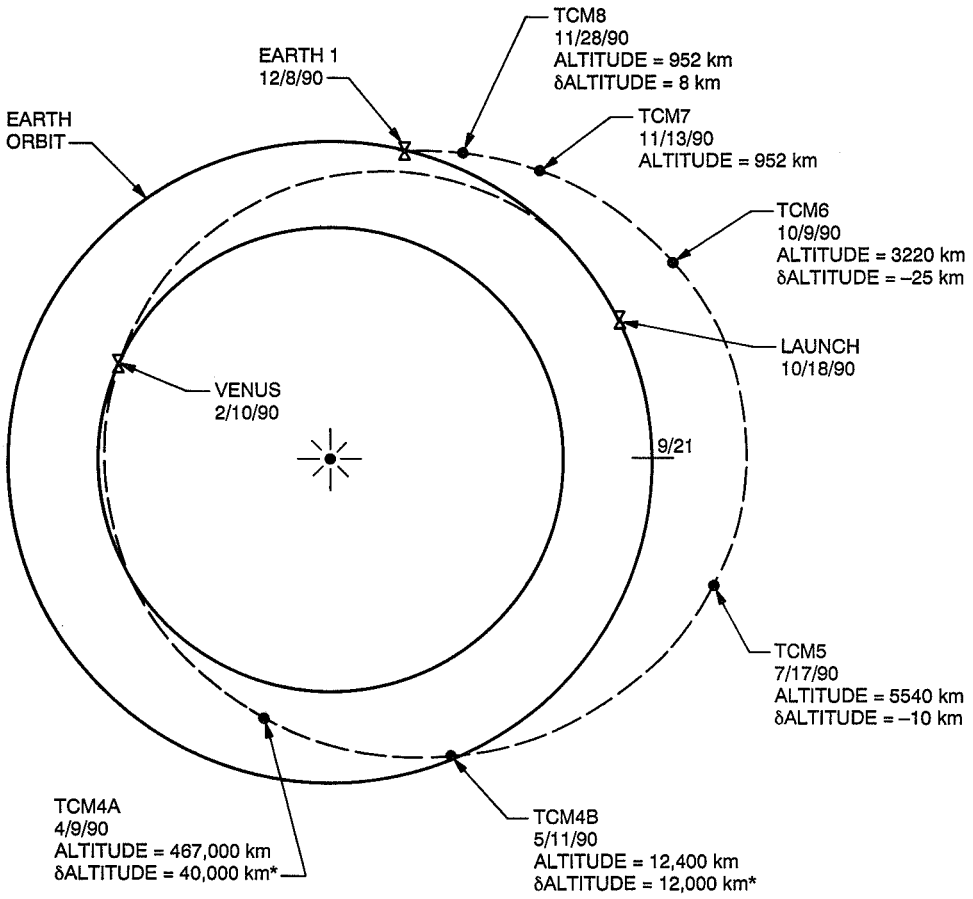


Fig. 1. ΔVLBI data flow.



*DELIBERATE SHORT BURN

Fig. 2. Galileo EGA 1 approach trajectory.

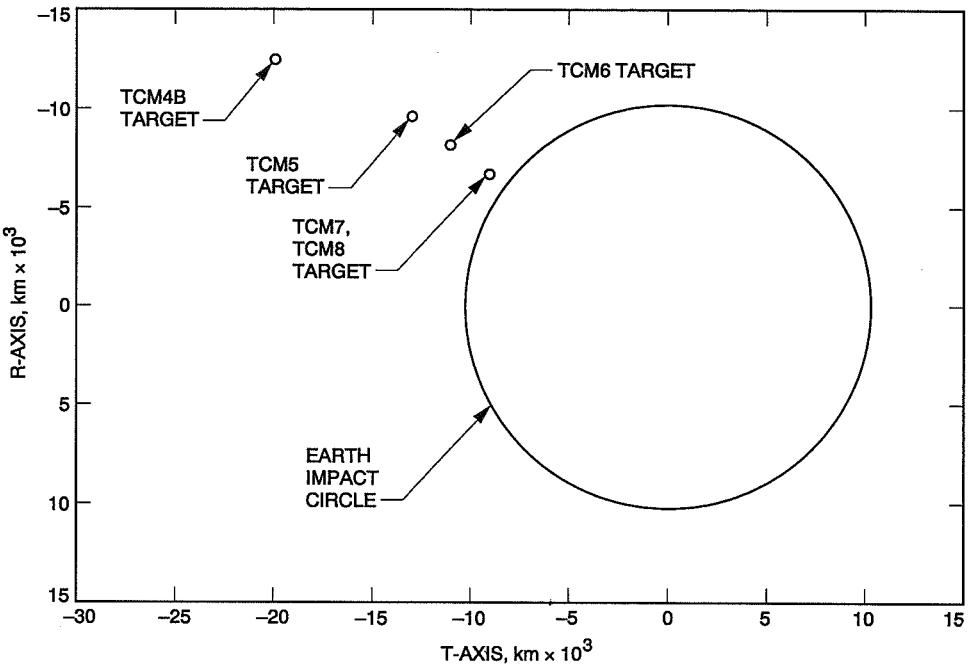


Fig. 3. B-plane aimpoints.

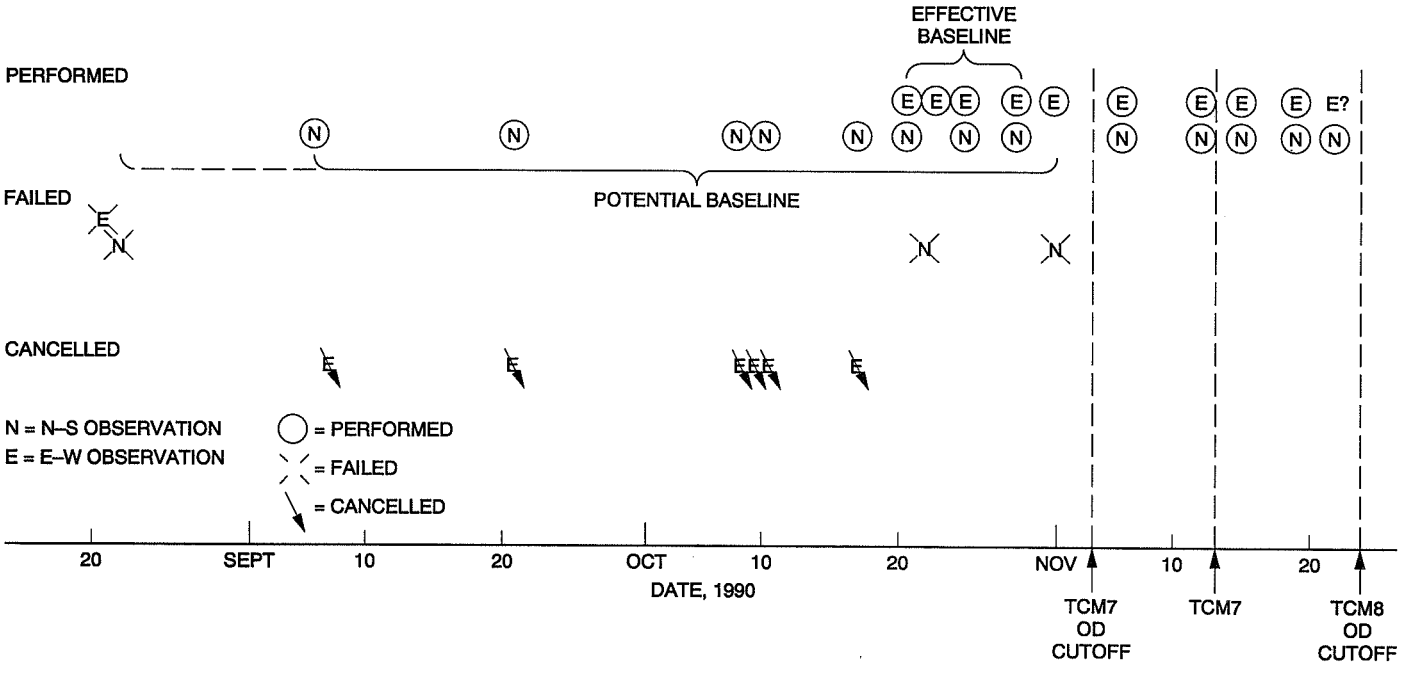


Fig. 4. ΔVLBI schedule.

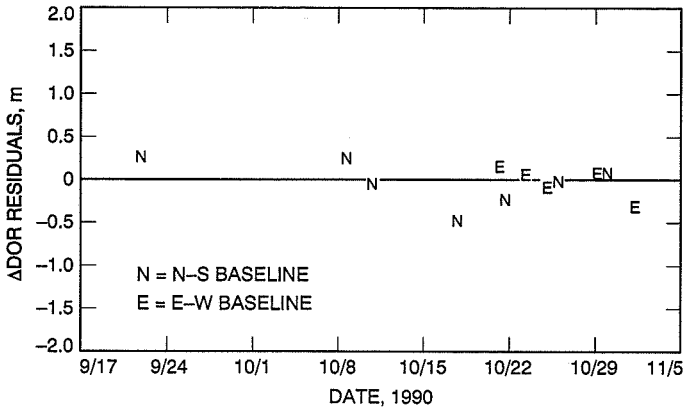


Fig. 5. ΔVLBI residuals for first cutoff.

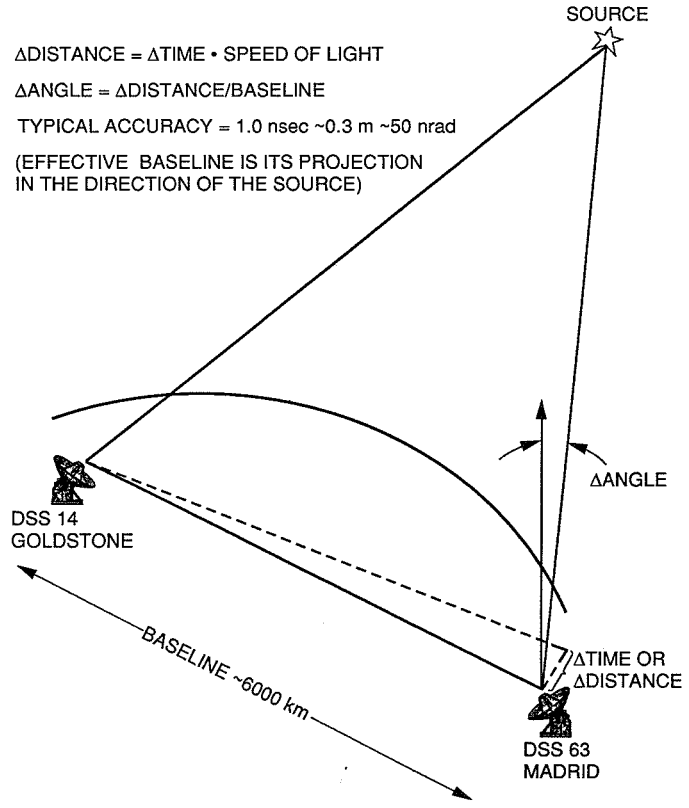


Fig. 6. Equivalent measures of ΔVLBI error.

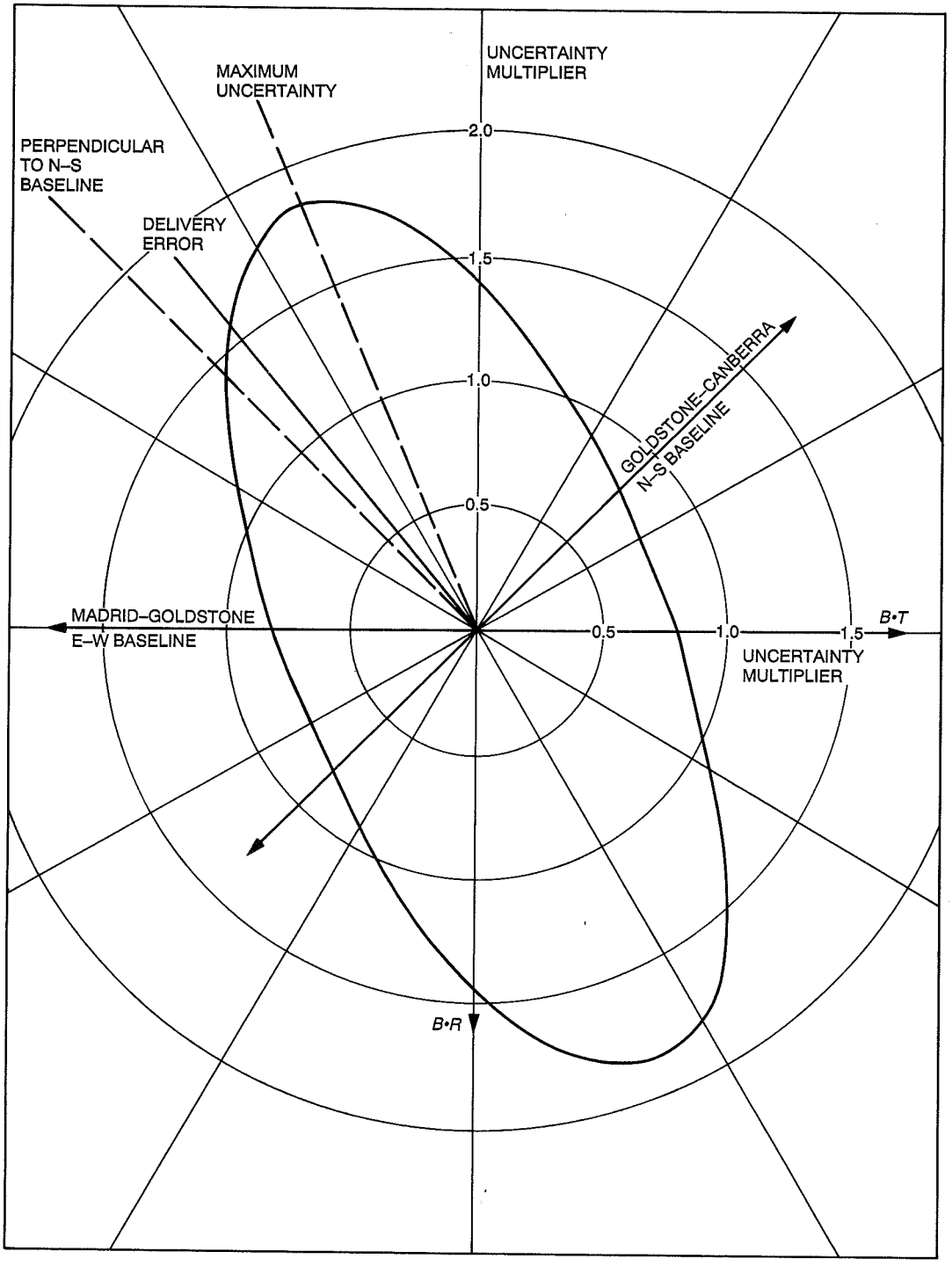


Fig. 7. Uncertainty multiplier for a ΔVLBI pair.

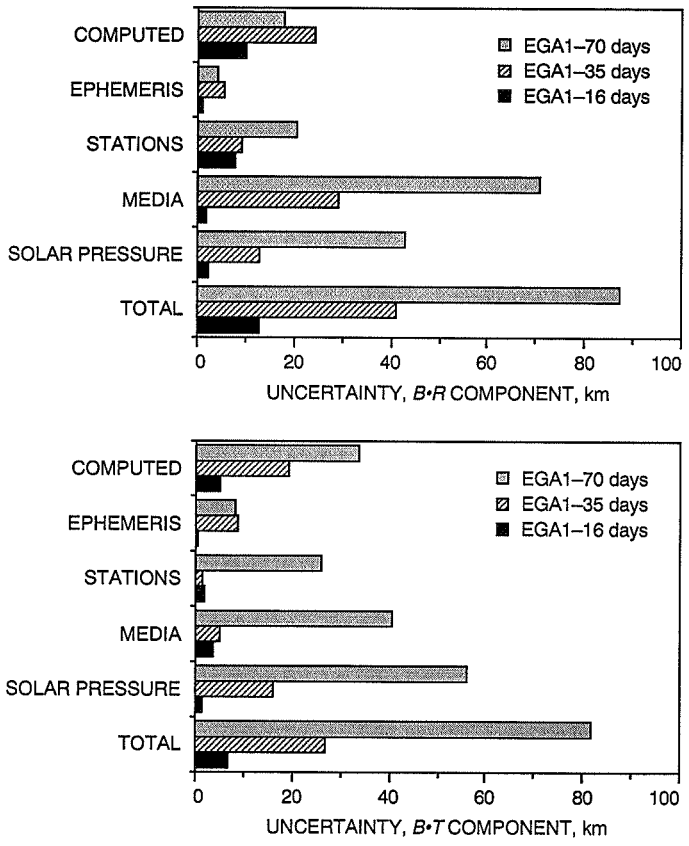


Fig. 8. Doppler data uncertainty (in the B-plane).

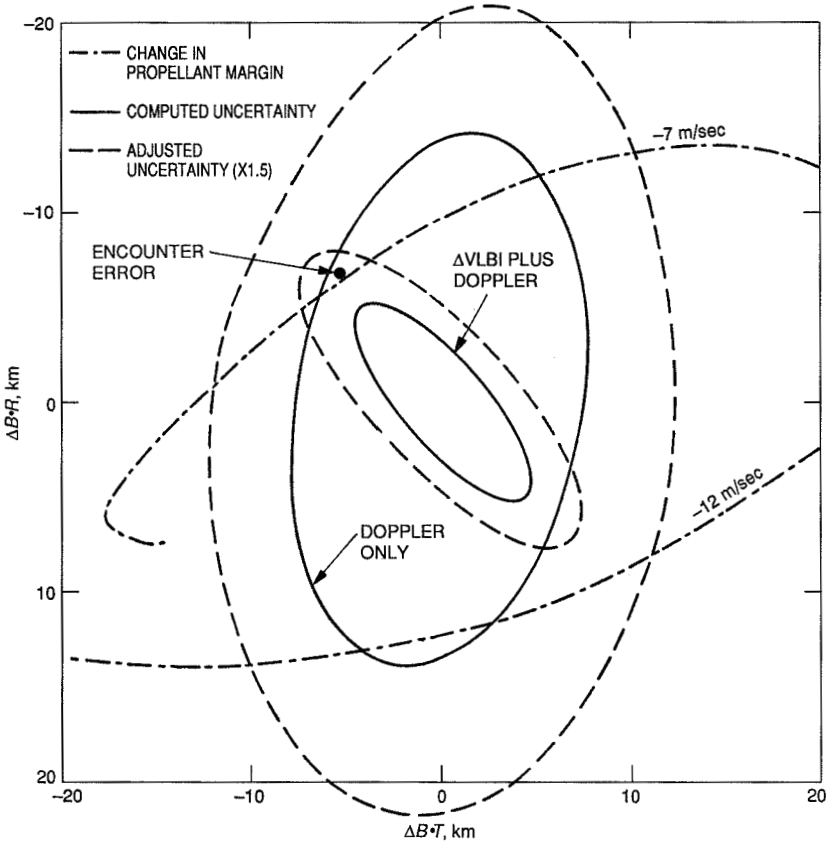
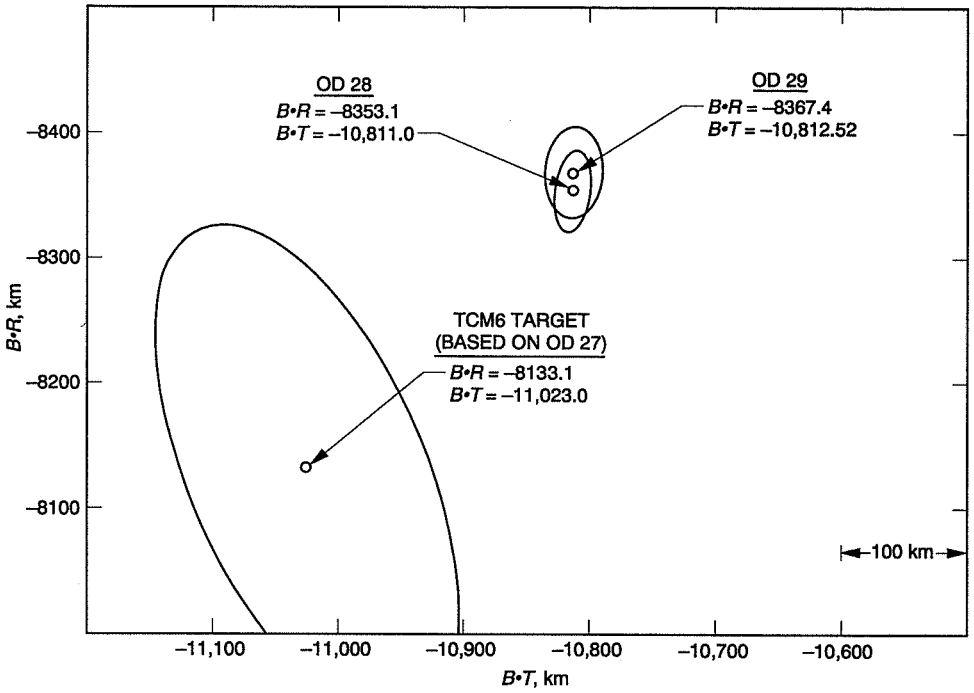


Fig. 9. B-plane delivery uncertainty and error.



DISCUSSION

- THE CURRENT B-PLANE SOLUTION IS ABOUT 300 km FROM OD 27 MAPPED USING THE BEST TCM6 ESTIMATE. WITH RESPECT TO STATISTICS DELIVERED WITH OD 27, THIS REPRESENTS A 3σ ERROR.
- SOLUTIONS BASED ON DOPPLER AND RANGE DATA ARE VERY SENSITIVE TO ERRORS IN NONGRAVITATIONAL FORCE MODELLING, ESPECIALLY SOLAR PRESSURE.
- REVISING THE EFFECTIVE SPACECRAFT AREA DOWNWARD, AND SOLVING FOR AN AXIAL ACCELERATION, REDUCES THE ERROR IN THE OD 27 SOLUTION.
- INCLUSION OF ΔDOR DATA IN THE OD SOLUTIONS REDUCES SENSITIVITY OF ESTIMATES TO NONGRAVITATIONAL MISMODELLING.

Fig. 10. Example and discussion of the Earth B-plane drift.

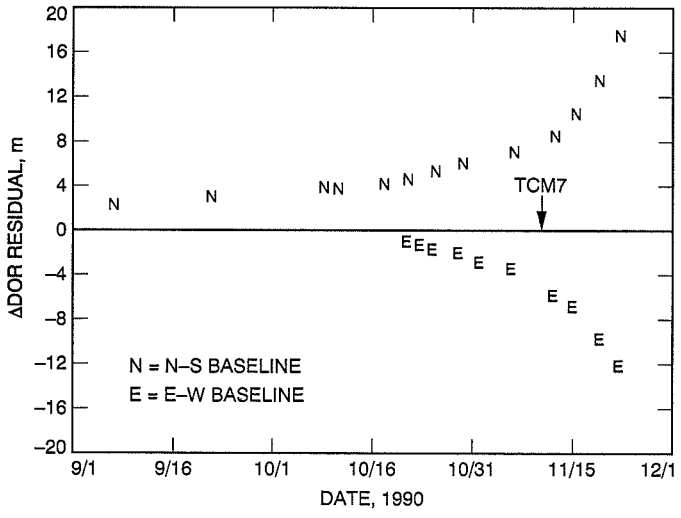


Fig. 11. ΔVLBI residuals for second cutoff of the Doppler-only solution.

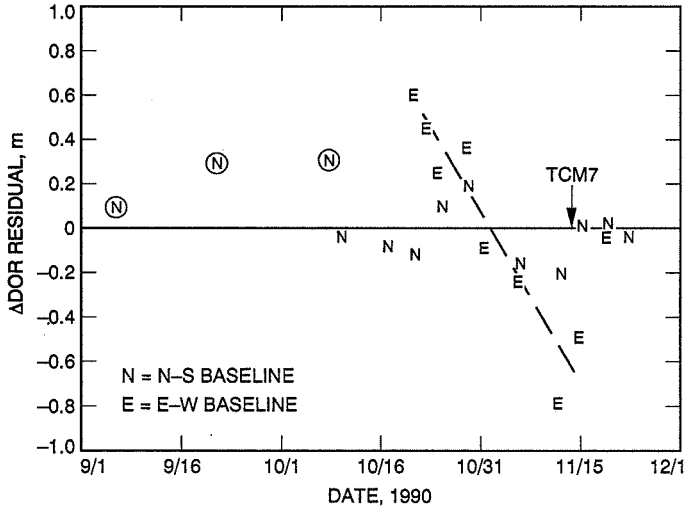


Fig. 12. ΔVLBI residuals for second cutoff of the Doppler-plus-ΔVLBI solution.

Appendix A

A Short History of the VLBI System and Its Improvements

The first ΔVLBI measurements of interplanetary spacecraft were made in the late 1970's by using radio astronomy equipment referred to as the Block 0 VLBI System. Data were recorded on videotape and shipped to a correlator facility for processing. The facility was located on the campus of the California Institute of Technology. Typical throughput time was four weeks. In 1981, the interim Block I VLBI System was developed by the DSN. This system consisted of computer-controlled data recording at the stations on nine-track tape, a satellite link to transmit the recorded data to the Jet Propulsion Laboratory, and a correlator at the Network Operations Control Center for data processing. Turnaround time as short as six hours was demonstrated, but constraints on scheduling data playbacks yielded a typical throughput time of 72 hours.

In 1986, the Narrow-Channel Bandwidth VLBI System replaced the interim Block I VLBI System to provide operational support for spacecraft navigation. Analog filters were replaced by digital filters. This change reduced key instrumental errors by a factor of two and improved overall system accuracy. The system configuration was automated and driven by station predicts (a computer data set consisting of the predicted antenna-pointing directions as a function of time). The option to record data on disk or tape was provided. The reliability of the system was improved in 1988 by developing a station-predicts–driven system for antenna pointing during ΔVLBI activities and by automating a procedure to detect and swap faulty channel downconverters. In 1989, the data playback rate was increased by a factor of two, reducing throughput time to less than 12 hours.

Appendix B

Example of a Detailed ΔVLBI Justification

Subject: Rationale for a ΔDOR Pair Two Days After TCM7

The Galileo Project Navigation Team has requested the Mission Control Team to negotiate tracking coverage to provide two ΔDOR activities after TCM7. If this negotiation is successful, it will mean TCM7 is "bracketed" by two pairs of baselines—the first one day before and another two days after. There are three reasons for this request.

First, the limited amount of time between the end of TCM7 and the TCM8 data cutoff (9 days) allows little time to develop a solution for plane-of-sky components for the maneuver. Since ΔDOR provides immediate information in the plane-of-sky, it follows that a pair of ΔDOR's after the maneuver will result in a better solution for TCM7 than can be obtained by Doppler.

Second, one may ask, Won't the ΔDOR pair scheduled on DOY 323 provide this plane-of-sky information? The answer is yes, it will, but this DOY 323 pair occurs three days after a SITURN (science turn) and retropropulsion module flushing. Any residual ΔV from the SITURN may appear as errors in TCM7. With no plane-of-sky measurement between TCM7 and the SITURN, the solution for TCM7 may become corrupted and that, in turn, will lead to a degraded B-plane solution.

A third reason for this request deals with redundancy for post-maneuver ΔDOR's. A ΔDOR activity is a complicated affair, involving spacecraft and ground configurations to be coordinated to the minute. The Navigation Team requires at least two successful baseline pairs following the TCM for any of the ΔDOR to be of use. With no baseline pair on DOY 319, a single point failure after TCM7 will endanger the success of the entire ΔDOR campaign.

In short, placement of a ΔDOR baseline pair two days after TCM7 will not only provide for a better OD solution, but will reduce the risk of suffering from the loss of one post-TCM7 ΔDOR.

Appendix C

Sample Calculation—Uncertainty of Estimation of Acceleration

To estimate an acceleration a by differencing two position measurements Δx_1 taken over a time interval Δt_1, one has

$$a = \frac{2(\Delta x_1)}{(\Delta t_1)^2}$$

The effect of that acceleration on a position at encounter Δx_e, propagated forward for a time interval Δt_2 from the last measurement, would be

$$\Delta x_e = \frac{(\Delta x_1)(\Delta t_2)^2}{(\Delta t_1)^2}$$

To obtain the effective uncertainty in Δx_1 requires taking the root sum square of the uncertainties at the ends of the time interval. In addition, the worst-case multiplier 1.85 from Fig. 7 will be used to obtain an upper bound (i.e., assuming an estimate is to be made in the most uncertain direction). Thus, using sig to denote the uncertainty,

$$\text{sig}(\Delta x_1) = 1.85\sqrt{(\text{sig}x_1)^2 + (\text{sig}x_2)^2}$$

Then, the resulting uncertainty in the encounter conditions is given by

$$\text{sig}(\Delta x_e) = \frac{\text{sig}(\Delta x_1)(\Delta t_2)^2}{(\Delta t_1)^2}$$

For the first example, consider ΔVLBI measurements taken at 90 days and 40 days before encounter, with the resulting acceleration estimate propagated the last 16 days to encounter. The inputs are

$$\Delta t_1 = 50 \text{ days}, \Delta t_2 = 16 \text{ days}$$

the measurement uncertainty is

$$\text{sig}(\Delta x_1) = 1.85\sqrt{1.17^2 + 2.90^2}$$

and the resulting encounter uncertainty is

$$\text{sig}(\Delta x_e) = 0.59 \text{ km}$$

Now, for comparison, let the first measurement be taken 48 days before encounter. The inputs are

$$\Delta t_1 = 8 \text{ days}, \Delta t_2 = 16 \text{ days}$$

the measurement uncertainty is

$$\text{sig}(\Delta x_1) = 1.85\sqrt{1.17^2 + 1.43^2}$$

and the resulting encounter uncertainty is

$$\text{sig}(x_e) = 13.7 \text{ km}$$

which is similar to the size of the actual encounter error in the B-plane. Thus, the accuracy is seen to be very dependent on the size of the effective measurement interval.

P.11

Carrier-to-Noise Power Estimation for the Block V Receiver

A. M. Monk
Telecommunications Systems Section

Two possible algorithms for the carrier-to-noise power (P_c/N_0) estimation in the Block V Receiver are analyzed and their performances compared. The expected value and the variance of each estimator algorithm are derived. The two algorithms examined here are known as the I-arm estimator, which relies on samples from only the in-phase arm of the digital phase-lock loop, and the IQ-arm estimator, which uses both in-phase and quadrature-phase arm signals. The IQ-arm algorithm is currently implemented in the Advanced Receiver II (ARX II). Both estimators are biased. The performance degradation due to phase jitter in the carrier tracking loop is taken into account. Curves of the expected value and the signal-to-noise ratio of the P_c/N_0 estimators versus actual P_c/N_0 are shown. From this, it is clear that the I-arm estimator performs better than the IQ-arm estimator when the data-to-noise power ratio (P_d/N_0) is high, i.e., at high P_c/N_0 values and a significant modulation index. When P_d/N_0 is low, the two estimators have essentially the same performance.

I. Introduction

In the Block V Receiver, P_c/N_0 estimates are made from accumulated samples in the carrier tracking loop. These estimates are used in various ways. Besides giving knowledge of the actual P_c/N_0 and as an indicator of when the loop is in-lock, the estimates can also be used in the Conscan process. (In Conscan, the pointing error of the Deep Space Network, DSN, ground antennas is reduced by conically scanning the antennas around a boresight. The present Conscan technique uses automatic gain control, AGC, samples of the ground receiver as a measure of the carrier power; but, if the noise power is fairly constant over several scan cycles, the P_c/N_0 estimates will serve the equivalent purpose.)

A simplified block diagram of the section of the digital phase-lock loop (DPLL) preceding the P_c/N_0 estimator is shown in Fig. 1. The sampling rate before the half-band filters (HBF's) is 40 MHz. The output of the HBF's has been decimated by two so that the new rate is one half of 40 MHz ($f_s = 1/T_s = 20$ MHz). The signals, i_n and q_n (shown below), are then accumulated over K samples, resulting in I_j and Q_j, which are the inputs to the P_c/N_0 estimator to be implemented in software. The sampling rate of I_j and Q_j is $f_s/K = 1/T_u$, where $T_u = KT_s$ is the carrier tracking loop update time, and can range from about 10 Hz to 10 kHz.

II. Estimator Algorithms

Assuming the carrier tracking loop is locked onto the received carrier frequency, the in-phase and quadrature-phase baseband signals (at sampling instant n) at the out-

put of the HBFs (neglecting higher order terms that are filtered out by the HBF's) are, respectively [1],

$$i_n = \sqrt{P_c}\cos\phi_n + \sqrt{P_d}d_n \mathrm{Sin}\left(\omega_{sc}nT_s + \theta_n\right)\sin\phi_n + n_i(n) \quad (1)$$

and

$$q_n = \sqrt{P_c}\sin\phi_n + \sqrt{P_d}d_n \mathrm{Sin}\left(\omega_{sc}nT_s + \theta_n\right)\cos\phi_n + n_q(n) \quad (2)$$

where P_c is the carrier power, ϕ_n is the phase error in the carrier tracking loop (assumed to be constant over one estimation period for all calculations), P_d is the data power, d_n represents the independent data samples that equal ± 1 with equal probability, and $\mathrm{Sin}(\omega_{sc}nT_s + \theta_n)$ represents a sample of a square-wave subcarrier with arbitrary phase θ_n.

In this case, $n_i(n)$ and $n_q(n)$ are the noise terms from the in-phase and quadrature-phase channels, respectively, and are assumed to be statistically independent Gaussian random variables with a mean of zero and variance of $\sigma^2 = N_0/(2T_s)$.

As seen in Fig. 1, the i_n and q_n signals are accumulated to give I_j and Q_j, respectively. Thus,

$$I_j = \frac{1}{K}\sum_{n=jK}^{(j+1)K-1}\sqrt{P_c}\cos\phi + n_I(j) \quad (3)$$

and

$$Q_j = \frac{1}{K}\sum_{n=jK}^{(j+1)K-1}\sqrt{P_d}d_n \mathrm{Sin}\left(\omega_{sc}nT_s + \theta_n\right)\cos\phi + n_Q(j) \quad (4)$$

where the $\sin\phi_n$ terms have been neglected under the assumption that the tracking error is very small (i.e., the loop is in lock). In addition, ϕ_n and θ_n have been replaced by ϕ and θ, respectively, under the assumption of constant phase over an observation interval. Both $n_I(j)$ and $n_Q(j)$ are formed by averaging samples of $n_i(n)$ and $n_q(n)$ and are, therefore, also independent Gaussian random variables with a mean of zero and a variance of $\sigma_I^2 = \sigma_Q^2 = N_0/(2KT_s) = N_0/(2T_u)$.

The IQ-arm estimator algorithm is [1]

$$\widehat{R_{IQ}} = \frac{\left(\frac{1}{L}\sum_{j=0}^{L-1} I_j\right)^2}{2T_u\left(\frac{1}{L}\sum_{j=0}^{L-1} Q_j^2\right)} \quad (5)$$

where the random variable $\widehat{R_{IQ}}$ is the P_c/N_0 estimator and L equals the number of DPLL updates in the estimation period.

By inserting Eqs. (3) and (4) into Eq. (5), it can be seen that the numerator of Eq. (5) essentially works to average out the zero mean noise samples $n_I(n)$, resulting in an estimate of the carrier power (assuming a small phase error). In the denominator of Eq. (5), the Q-arm estimates are first squared, then averaged. Assuming that the subcarrier frequency ω_{sc} is large as compared with $1/T_s$ and the data power is not too high, the first term in Eq. (4) will be small and, thus, the denominator of Eq. (5) will be a scaled estimate of the sum of the squares of $n_Q(n)$, which equals the noise power. (In reality, the effect of the data power is important, as will be seen in the analysis below.)

The I-arm estimator algorithm [2] is

$$\widehat{R_I} = \frac{\left(\frac{1}{L}\sum_{j=0}^{L-1} I_j\right)^2}{\frac{2T_u}{L-1}\left[\sum_{j=0}^{L-1} I_j^2 - \frac{1}{L}\left(\sum_{j=0}^{L-1} I_j\right)^2\right]} \quad (6)$$

where $\widehat{R_I}$ is the P_c/N_0 estimator and L is defined as above.

By inserting Eq. (3) into Eq. (6), it can be seen that the numerator of Eq. (6) is the same as that of Eq. (5). However, the denominator, which reduces to an estimate of noise power, does not depend on the data power and, as shall be seen below, is the major advantage of the I-arm estimator over the IQ-arm estimator.

III. First and Second Moments of the IQ- and I-Arm Estimators

For simplicity, the following intermediate variables are defined:

$$X = \left(\frac{1}{L} \sum_{j=0}^{L-1} I_j \right)^2 \tag{7}$$

$$Y = \frac{1}{L} \sum_{j=0}^{L-1} Q_j^2 \tag{8}$$

$$Z = \sum_{j=0}^{L-1} I_j^2 \tag{9}$$

Using Eqs. (7), (8), and (9), Eqs. (5) and (6) can be written as

$$\widehat{R_{IQ}} = \frac{X}{2T_u Y} \tag{10}$$

and

$$\widehat{R_I} = \frac{X}{\frac{2T_u}{L-1}(Z - LX)} \tag{11}$$

The statistics of the estimators are not trivial, since in both cases the numerator and the denominator are correlated. (In the case of the IQ-arm estimator, the numerator and denominator are correlated, even though they are obtained from in-phase and quadrature-phase arms, since both arms have terms that contain the phase error ϕ.)

The method used here to find the statistics is to expand the expressions for $\widehat{R_{IQ}}$ and $\widehat{R_I}$ (each of which is a function of two random variables whose moments are shown below) in a two-dimensional Taylor series and then to find the expected value and variance of the expanded series. The higher order terms of the Taylor series expansion are ignored. Using this approach, the expected value of the IQ-arm estimator can be approximated by the following expression [3,4]:

$$\overline{\widehat{R_{IQ}}} = \widehat{R_{IQ}}\Big|_{\overline{X},\overline{Y}} + \frac{1}{2}\left[\frac{\partial^2 \widehat{R_{IQ}}}{\partial X^2} \mathrm{var}(X) + \frac{\partial^2 \widehat{R_{IQ}}}{\partial Y^2} \mathrm{var}(Y) \right]\Bigg|_{\overline{X},\overline{Y}} + \frac{\partial^2 \widehat{R_{IQ}}}{\partial X \partial Y} \mathrm{cov}(X,Y)\Big|_{\overline{X},\overline{Y}} \tag{12}$$

and the variance by

$$\mathrm{var}\left(\widehat{R_{IQ}}\right) = \left(\frac{\partial \widehat{R_{IQ}}}{\partial X} \right)^2\Bigg|_{\overline{X},\overline{Y}} \mathrm{var}(X) + \left(\frac{\partial \widehat{R_{IQ}}}{\partial Y} \right)^2\Bigg|_{\overline{X},\overline{Y}} \mathrm{var}(Y) + 2\, \frac{\partial \widehat{R_{IQ}}}{\partial X} \frac{\partial \widehat{R_{IQ}}}{\partial Y} \mathrm{cov}(X,Y)\Big|_{\overline{X},\overline{Y}} \tag{13}$$

where the overbar indicates the statistical average and the covariance of X and Y is defined as $\mathrm{cov}(X,Y) \triangleq \mathbf{E}\{(X - \overline{X})(Y - \overline{Y}\} = \overline{XY} - \overline{X}\,\overline{Y}$. The expected value operator is represented by $\mathbf{E}\{\cdot\}$. For the I-arm estimator, the equations are identical, except that Y is replaced with Z.

The moments of X, Y, and Z, which are necessary for the computation of Eqs. (12) and (13), are shown here (see the Appendix for the derivations).

Defining $g_2 \triangleq \overline{\cos^2 \phi}$ and $g_4 \triangleq \overline{\cos^4 \phi}$, the first-order moments are given by

$$\overline{X} = g_2 P_c + \frac{N_0}{2LT_u} \tag{14}$$

$$\overline{Y} = K^{-1} g_2 P_d + \frac{N_0}{2T_u} \tag{15}$$

$$\overline{Z} = L g_2 P_c + \frac{LN_0}{2T_u} \tag{16}$$

Eq. (14) shows that the numerators of both estimators, Eqs. (5) and (6), are, in fact, a positively biased estimate of the carrier power (neglecting phase error). Also, the estimate of noise power in the IQ-arm case, given by a

scaled version of Eq. (15), is biased. The I-arm estimate of the noise power, however, is unbiased.

The second-order moments are given by

$$\overline{X^2} = g_4 P_c^2 + \frac{6N_0}{2LT_u} g_2 P_c + 3\left(\frac{N_0}{2LT_u}\right)^2 \tag{17}$$

$$\overline{Y^2} = \frac{3K-2}{K^3} g_4 P_d^2 + \frac{L+2}{LKT_u} N_0 g_2 P_d + \frac{L+2}{L}\left(\frac{N_0}{2T_u}\right)^2 \tag{18}$$

$$\overline{Z^2} = L^2 g_4 P_c^2 + L(L+2)\frac{N_0}{T_u} g_2 P_c + L(L+2)\left(\frac{N_0}{2T_u}\right)^2 \tag{19}$$

and the covariances by

$$\operatorname{cov}(X,Y) = K^{-1}\left(g_4 - g_2^2\right) P_d P_c \tag{20}$$

$$\operatorname{cov}(X,Z) = L\left(g_4 - g_2^2\right) P_c^2 + 2T_u^{-1} N_0 g_2 P_c + \frac{N_0^2}{2LT_u^2} \tag{21}$$

In the case of the Block V receiver, the DPLL is a third-order loop whose phase error has a probability density function characterized, approximately, by a Tikhonov distribution (for zero detuning) [5]. For this case, g_2 and g_4 are obtained by using the relation $\overline{\cos n\phi} = [I_n(\alpha)]/[I_0(\alpha)]$ [5] where, for high loop signal-to-noise ratio ($\rho = P_c/N_0 B_L$), α is approximated by ρ [5]. Thus,

$$\left.\begin{aligned} \overline{\cos^2\phi} &= \frac{1}{2} + \frac{I_2(\rho)}{2I_0(\rho)} \\ \overline{\cos^4\phi} &= \frac{1}{8}\left[3 + \frac{4I_2(\rho)}{I_0(\rho)} + \frac{I_4(\rho)}{I_0(\rho)}\right] \end{aligned}\right\} \tag{22}$$

By inserting Eqs. (14) through (22) into Eqs. (12) and (13), the mean for the IQ-arm estimator, after simplifying, is

$$\overline{\widehat{R_{IQ}}} = \frac{KA}{2LT_u B} + \frac{2T_u KA}{LB^3}\left\{\left[\left(3 - \frac{2}{K}\right) g_4 - g_2^2\right]\left(\frac{P_d}{N_0}\right)^2 + \frac{2Kg_2}{LT_u}\frac{P_d}{N_0} + \frac{K^2}{2LT_u^2}\right\} - \frac{2KT_u}{B^2}\left(g_4 - g_2^2\right)\frac{P_d}{N_0}\frac{P_c}{N_0} \tag{23}$$

and the variance is

$$\operatorname{var}\left(\widehat{R_{IQ}}\right) = \frac{K^2}{B^2}\left\{\left(g_4 - g_2^2\right)\left(\frac{P_c}{N_0}\right)^2 + \frac{2g_2}{LT_u}\frac{P_c}{N_0} + \frac{1}{2L^2T_u^2}\right\} + \frac{K^2A^2}{L^2B^4}\left\{\left[\left(3 - \frac{2}{K}\right) g_4 - g_2^2\right]\left(\frac{P_d}{N_0}\right)^2 + \frac{2Kg_2}{LT_u}\frac{P_d}{N_0} + \frac{K^2}{2LT_u^2}\right\} - \frac{2K^2A}{LB^3}\left(g_4 - g_2^2\right)\frac{P_d}{N_0}\frac{P_c}{N_0} \tag{24}$$

where $A = 2LT_u g_2(P_c/N_0)+1$ and $B = 2T_u g_2(P_d/N_0)+K$.

The mean for the I-arm estimator is

$$\overline{\widehat{R_I}} = \frac{L+1}{L-1}\left(g_2\frac{P_c}{N_0} + \frac{1}{2LT_u}\right) \tag{25}$$

and the variance is

$$\operatorname{var}\left(\widehat{R_I}\right) = \left(g_4 - \frac{L-3}{L-1} g_2^2\right)\left(\frac{P_c}{N_0}\right)^2 + \frac{2g_2}{T_u(L-1)}\frac{P_c}{N_0} + \frac{1}{2T_u^2 L(L-1)} \tag{26}$$

The signal-to-noise ratio (SNR) of the estimators is defined here as

$$SNR\left(\hat{R}\right) = \frac{\overline{\hat{R}}^2}{\operatorname{var}\left(\hat{R}\right)} \tag{27}$$

It should be noted that for P_d/N_0 equal to zero and $L \gg 1$, the IQ-arm estimators expected value and variance reduce almost exactly to that of the I-arm estimator.

IV. Discussion of the Results

The performance of the estimators is shown in Figs. (2) and (3). Figure 2 shows $\overline{\widehat{R_{IQ}}}$ and SNR ($\widehat{R_{IQ}}$) versus the carrier-to-noise power (P_c/N_0) in dB-Hz for a typical modulation index such as that used by the Galileo spacecraft,

C-5

i.e., $\Delta = 76$ deg where $\tan^2 \Delta = P_d/P_c$. Figure 3 depicts $\widehat{\overline{R_I}}$ and SNR $(\widehat{R_I})$ versus P_c/N_0 in dB-Hz. The effect of the data power P_d in the IQ-arm estimator becomes critical when the data power is large, which, for a $\Delta = 76$ deg, occurs at high values of P_c/N_0. At this point, the mean of the IQ-arm estimate begins to diverge rapidly from the true value, in contrast to the mean of the I-arm estimate, which asymptotically approaches the true value. (This contrast is more clearly seen when comparing the SNR curves.) *At low* P_c/N_0 *the I- and IQ-arms are essentially identical.* The expected values of both estimators approach a positive asymptote at low values of P_c/N_0 as a result of the positive bias in the estimator, which decreases with increasing L. Also, the effect of the loop bandwidth B_L on both estimators can be seen. This is the result of tracking loop jitter at low P_c/N_0 caused by the g_2 and g_4 terms in Eqs. (23) through (26). (When B_L gets larger, the phase error variance increases causing g_2 and g_4 to decrease, thus, biasing the estimates.)

The I-arm estimator SNR approaches an asymptotic value at high P_c/N_0 values as a result of the bias in the estimator. This asymptotic value increases proportionately with the estimation length, L. (For $P_c/N_0 \gg 1$ and $L \gg 1$, $\mathrm{SNR}(\widehat{R_I}) \rightarrow L/2$). To illustrate the meaning of the SNR in Fig. 3, consider the point $P_c/N_0 = 40$ dB-Hz where $\mathrm{SNR}(\hat{R}) = 24$ dB. This means that the expected value of $\hat{R}$ is 12 dB above the standard deviation of $\hat{R}$ (or $\sigma_{\hat{R}} = 10^{-1.2}\overline{\hat{R}}$). Thus, the tolerance on the estimator reading is $10,000\ (1 \pm 10^{-1.2})$, which is approximately equal to 40 ± 0.27 dB.

V. Conclusions

In this article, the expected value and variance of two P_c/N_0 estimators were derived and plotted. The results show that the I-arm estimator performs better than the IQ-arm estimator when P_d/N_0 is high, which occurs for high values of P_c/N_0 and significant modulation index. (In Fig. 2, this happens when P_c/N_0 exceeds 38 dB-Hz.) Also the effect of the positive estimator bias *and* loop jitter, which increases for higher values of B_L, is to degrade the performance of both estimators at low values of P_c/N_0.

Acknowledgments

The author thanks Sami Hinedi, Alex Mileant, Ted Peng, Ramin Sadr, Biren Shah, and Marvin Simon for their many helpful comments and suggestions.

References

[1] S. Hinedi, "A Functional Description of the Advanced Receiver," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 131–149, February 15, 1989.

[2] S. Hinedi and J. I. Statman, *High-Dynamic GPS Tracking*, JPL Publication 88-35, Jet Propulsion Laboratory, Pasadena, California, December 15, 1989.

[3] M. K. Simon and A. Mileant, "SNR Estimation for the Baseband Assembly," *TDA Progress Report 42-85*, vol. January–March 1986, Jet Propulsion Laboratory, Pasadena, California, pp. 118–126, February 15, 1986.

[4] M. G. Kindal and A. Stuart, *The Advanced Theory of Statistics, vol. I*, New York: Hafner Publishing Co., 1973.

[5] W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering*, New Jersey: Prentice Hall, Inc., pp. 31–36, 1973.

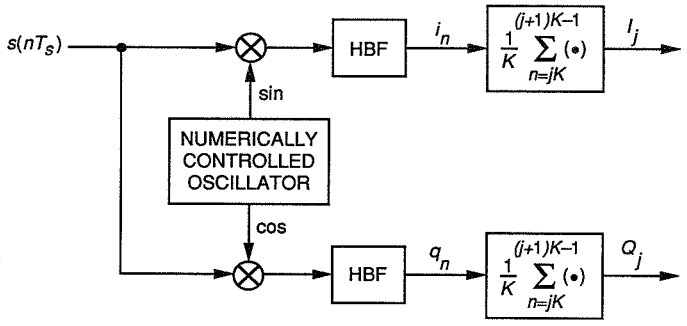


Fig. 1. Part of the DPLL preceding the P_c/N_0 estimator.

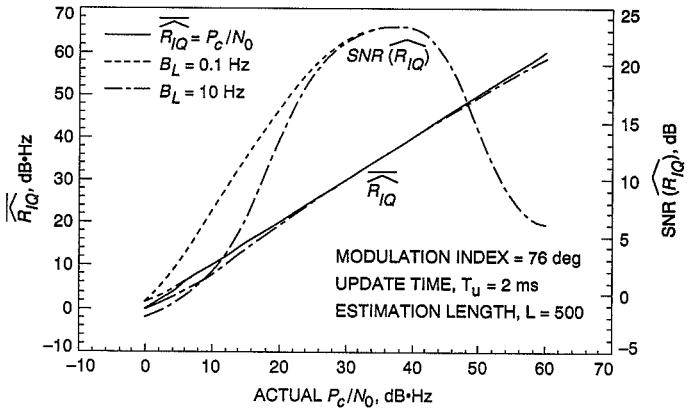


Fig. 2. Mean value and SNR of the IQ-arm estimator.

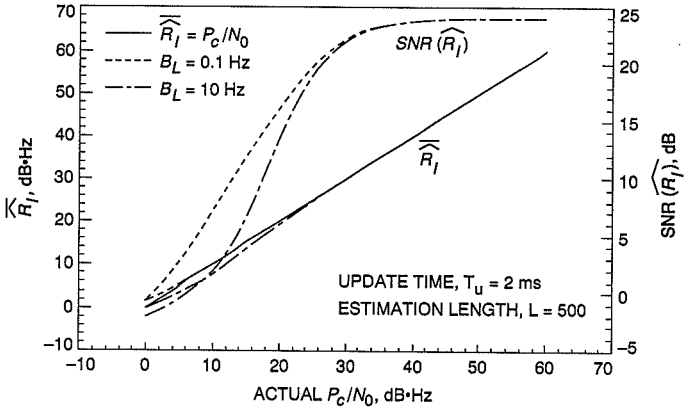


Fig. 3. Mean value and SNR of the I-arm estimator.

Appendix
Derivation of Equations

For ease of notation, define $g_n \triangleq \overline{\cos^n \phi}$.

I. Evaluation of the First Moments

The first moments given in Eqs. (14) through (16) are derived, using Eqs. (3) and (7), as follows:

$$\overline{X} = \mathbf{E}\left\{\frac{1}{L^2}\left[\sum_{j=0}^{LK-1} K^{-1}\sqrt{P_c}\cos\phi + \sum_{j=0}^{L-1} n_I(j)\right]^2\right\}$$

$$= \mathbf{E}\left\{P_c\cos^2\phi + \frac{2}{L}\sqrt{P_c}\cos\phi\sum_{j=0}^{L-1} n_I(j) + \frac{1}{L^2}\left[\sum_{j=0}^{L-1} n_I(j)\right]^2\right\}$$

$$= P_c g_2 + \frac{N_0}{2LT_u} \tag{A-1}$$

The expression for $\overline{Y}$ contains products of d_n, $\mathsf{S}\text{in}\,(\omega_{sc}nT_s + \theta)$ and $\cos\phi$. It is assumed that each of these terms is independent of the others. Here, $\mathsf{S}\text{in}\,(\omega_{sc}nT_s + \theta)$ is represented by S_n (recall this is the square-wave subcarrier term). Thus, for example,

$$\mathbf{E}\left\{d_n d_m \mathsf{S}_n S_m \cos^2\phi\right\} = \delta_{nm} g_2 \tag{A-2}$$

since $\mathsf{S}_n\mathsf{S}_n = 1$ and $\mathbf{E}\{d_n d_m\} = \delta_{nm}$, where δ_{nm} is the two-dimensional Kronecker delta function.

Using Eqs. (4) and (8) yields

$$\overline{Y} = \mathbf{E}\left\{\frac{1}{L}\sum_{j=0}^{L-1}\left[\frac{P_d}{K^2}\sum_{n=jK}^{(j+1)K-1}\sum_{m=jK}^{(j+1)K-1} d_n d_m \mathsf{S}_n S_m \cos^2\phi + \frac{2\sqrt{P_d}}{K}\sum_{n=jK}^{(j+1)K-1} d_n \mathsf{S}_n \cos\phi n_Q(j) + n_Q^2(j)\right]\right\}$$

$$= \frac{P_d}{K} g_2 + \frac{1}{L}\sum_{j=0}^{L-1}\mathbf{E}\left\{n_Q^2(j)\right\}$$

$$= \frac{P_d}{K} g_2 + \frac{N_0}{2T_u} \tag{A-3}$$

Using Eqs. (3) and (9) yields

$$\overline{Z} = \mathbf{E}\left\{\sum_{j=0}^{L-1}\left[\frac{1}{K}\sum_{n=jK}^{(j+1)K-1}\sqrt{P_c}\cos\phi + n_I(j)\right]^2\right\}$$

$$= \mathbf{E}\left\{LP_c\cos\phi + 2L\sqrt{P_c}\cos\phi\sum_{j=0}^{L-1}n_I(j) + \sum_{j=0}^{L-1}n_I^2(j)\right\}$$

$$= LP_c g_2 + \frac{LN_0}{2T_u} \tag{A-4}$$

II. Derivation of Second Moments

The second moment of X, using Eqs. (3) and (7), is:

$$\overline{X^2} = \mathbf{E}\left\{\frac{1}{L^4}\left[\sum_{j=0}^{LK-1}K^{-1}\sqrt{P_c}\cos\phi + \sum_{j=0}^{L-1}n_I(j)\right]^4\right\}$$

$$= g_4P_c^2 + 6P_cg_2\mathbf{E}\left\{\left[\frac{1}{L}\sum_{j=0}^{L-1}n_I(j)\right]^2\right\} + \mathbf{E}\left\{\frac{1}{L}\left[\sum_{j=0}^{L-1}n_I(j)\right]^4\right\}$$

$$= g_4P_c + \frac{6N_0}{2LT_u}g_2P_c + 3\left(\frac{N_0}{2LT_u}\right)^2 \tag{A-5}$$

Equations (A-6) through (A-11) are needed to compute $\mathbf{E}\{Y^2\}$.

$$\mathbf{E}\{d_n\} = 0 \tag{A-6}$$

$$\mathbf{E}\{d_nd_md_pd_r\} = \begin{cases} 1 \text{ if } \begin{cases} n=m\neq p=r & \text{or} \\ n=p\neq m=r & \text{or} \\ n=r\neq m=r & \text{or} \\ n=m=p=r \end{cases} \\ 0 \text{ otherwise} \end{cases} \tag{A-7}$$

or, in a more mathematical form,

$$\mathbf{E}\{d_nd_md_pd_r\} = (\delta_{nm}\delta_{pr} - \delta_{nmpr}) + (\delta_{np}\delta_{mr} - \delta_{nmpr}) + (\delta_{nr}\delta_{mr} - \delta_{nmpr}) + \delta_{nmpr}$$

$$= \delta_{nm}\delta_{pr} + \delta_{np}\delta_{mr} + \delta_{nr}\delta_{mr} - 2\delta_{nmpr} \tag{A-8}$$

where δ_{nmpr} is the four-dimensional Kronecker delta function, which equals unity only when $n = m = p = r$.

Using Eq. (A-8),

$$\sum_{n=jK}^{(j+1)K-1}\sum_{m=jK}^{(j+1)K-1}\sum_{n=jK}^{(j+1)K-1}\sum_{n=jK}^{(j+1)K-1}\mathbf{E}\{d_n d_m d_p d_r\} = K^2 + K^2 + K^2 - 2K$$

$$= 3K^2 - 2K \tag{A-9}$$

If n_i represents independent, Gaussian, zero mean random variables, then

$$\mathbf{E}\{n_1 n_2 n_3 n_4\} = \mathbf{E}\{n_1 n_2\}\mathbf{E}\{n_3 n_4\} + \mathbf{E}\{n_1 n_3\}\mathbf{E}\{n_2 n_4\} + \mathbf{E}\{n_1 n_4\}\mathbf{E}\{n_2 n_3\} \tag{A-10}$$

Using Eq. (A-10),

$$\mathbf{E}\left\{\sum_{j=0}^{L-1}\sum_{l=0}^{L-1} n_Q^2(j) n_Q^2(l)\right\} = \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\mathbf{E}\{n_Q^2(j) n_Q^2(l)\}$$

$$= \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\mathbf{E}\{n_Q^2(j)\}\mathbf{E}\{n_Q^2(l)\} + 2\left[\mathbf{E}\{n_Q(j) n_Q(l)\}\right]^2$$

$$= \sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\left(\sigma_I^4 + 2\left[\delta_{jl}\sigma_I^2\right]^2\right)$$

$$= \left(L^2 + 2L\right)\sigma_I^4 \tag{A-11}$$

Using Eqs. (4), (8), (A-9), and (A-11) yields

$$\overline{Y^2} = \mathbf{E}\left\{\frac{1}{L^2}\sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\left\{\left[\frac{P_d}{K^2}\sum_{n=jK}^{(j+1)K-1}\sum_{m=jK}^{(j+1)K-1} d_n d_m \eth_n S_m \cos^2\phi + \frac{2\sqrt{P_d}}{K}\sum_{n=jK}^{(j+1)K-1} d_n \eth_n \cos\phi n_Q(j) + n_Q^2(j)\right]\right.\right.$$

$$\left.\left.\times\left[\frac{P_d}{K^2}\sum_{p=lK}^{(l+1)K-1}\sum_{r=lK}^{(l+1)K-1} d_p d_r S_p S_r \cos^2\phi + \frac{2\sqrt{P}_d}{K}\sum_{p=lK}^{(l+1)K-1} d_p S_p \cos\phi n_Q(l) + n_Q^2(l)\right]\right\}\right\}$$

$$= \frac{P_d^2}{K^4}(3K^2 - 2K)g_4 + 2\left(\frac{P_d}{K^2}K g_2\frac{N_0}{2T_u}\right) + \frac{4}{K^2}P_d\frac{K}{L}g_2\frac{N_0}{2T_u} + \frac{1}{L^2}\left(L^2 + 2L\right)\left(\frac{N_0}{2T_u}\right)^2$$

$$= \frac{3K-2}{K^3}g_4 P_d^2 + \frac{L+2}{LK}g_2\frac{N_0}{T_u}P_d + \frac{L+2}{L}\left(\frac{N_0}{2T_u}\right)^2 \tag{A-12}$$

The second moment of Z is given, from Eqs. (3) and (9), by

$$\overline{Z^2} = \mathbf{E}\left\{\sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\left[\frac{1}{K}\sum_{n=jK}^{(j+1)K-1}\sqrt{P_c}\cos\phi + n_I(j)\right]^2\left[\frac{1}{K}\sum_{m=lK}^{(l+1)K-1}\sqrt{P_c}\cos\phi + n_I(l)\right]^2\right\}$$

$$= \mathbf{E}\left\{\sum_{j=0}^{L-1}\sum_{l=0}^{L-1}\left[P_c\cos^2\phi + 2\sqrt{P_c}\cos\phi\, n_I(j) + n_I^2(j)\right]\left[P_c\cos^2\phi + 2\sqrt{P_c}\cos\phi\, n_I(l) + n_I^2(l)\right]\right\}$$

$$= L^2P_c^2g_4 + 2\left(L^2P_cg_2\frac{N_0}{2T_u}\right) + 4LP_cg_2\frac{N_0}{2T_u} + (L^2+2L)\left(\frac{N_0}{2T_u}\right)^2 \tag{A-13}$$

The cross-correlation of X and Z is found as follows, from Eqs. (3), (7), and (9):

$$XZ = \frac{1}{L^2}\left(\sum_{j=0}^{LK-1}K^{-1}\sqrt{P_c}\cos\phi + \sum_{j=0}^{L-1}n_I(j)\right)^2\sum_{j=0}^{L-1}\left[\frac{1}{K}\sum_{n=jK}^{(j+1)K-1}\sqrt{P_c}\cos\phi + n_I(j)\right]^2$$

$$= \left\{P_c\cos^2\phi + \frac{2}{L}\sqrt{P_c}\cos\phi\sum_{j=0}^{L-1}n_I(j) + \frac{1}{L^2}\left[\sum_{j=0}^{L-1}n_I(j)\right]^2\right\}$$

$$\times\left[LP_c\cos^2\phi + 2L\sqrt{P_c}\cos\phi\sum_{j=0}^{L-1}n_I(j) + \sum_{j=0}^{L-1}n_I^2(j)\right] \tag{A-14}$$

Using Eq. (A-10), the expectation of the fourth-order noise cross-terms in Eq. (A-14) is

$$\mathbf{E}\left\{\left(\sum_{j=0}^{L-1}n_I(j)\right)^2\sum_{j=0}^{L-1}n_I^2(j)\right\} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\sum_{k=0}^{L-1}\mathbf{E}\left\{n_I(i)n_I(j)n_I^2(k)\right\}$$

$$= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\sum_{k=0}^{L-1}\mathbf{E}\left\{n_I(i)n_I(j)\right\}\mathbf{E}\left\{n_I^2(k)\right\} + 2\mathbf{E}\left\{n_I(i)n_I(k)\right\}\mathbf{E}\left\{n_I(j)n_I(k)\right\}$$

$$= \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\sum_{k=0}^{L-1}\delta_{ij}\sigma_I^2\sigma_I^2 + 2\delta_{ik}\sigma_I^2\delta_{jk}\sigma_I^2$$

$$= L^2\sigma_I^4 + 2L\sigma_I^4 \tag{A-15}$$

where $\sigma_I^2 = N_0/2T_u$.

Thus, using Eqs. (A-14) and (A-15),

$$\overline{XZ} = LP_c^2 g_4 + LP_c g_2 \frac{N_0}{2T_u} + \frac{4}{L} P_c g_2 \frac{LN_0}{2T_u} + \frac{1}{L} P_c g_2 \frac{LN_0}{2T_u} + \frac{1}{L^2} \mathbf{E} \left\{ \left[\sum_{j=0}^{L-1} n_I(j) \right]^2 \sum_{j=0}^{L-1} n_I^2(j) \right\}$$

$$= LP_c^2 g_4 + (L+5) P_c g_2 \frac{N_0}{2T_u} + \left(1 + \frac{2}{L}\right) \left(\frac{N_0}{2T_u}\right)^2 \tag{A-16}$$

The cross-correlation of X and Y is found in a similar way, by making use of Eqs. (3), (4), (7), (8), (A-2), and (A-10),

$$\overline{XY} = \mathbf{E} \left\{ \frac{1}{L^2} \left(\sum_{j=0}^{LK-1} K^{-1} \sqrt{P_c} \cos\phi + \sum_{j=0}^{L-1} n_I(j) \right)^2 \right.$$

$$\left. \times \frac{1}{L} \sum_{j=0}^{L-1} \left[\frac{P_d}{K^2} \sum_{n=jK}^{(j+1)K-1} \sum_{m=jK}^{(j+1)K-1} d_n d_m \Im_n S_m \cos^2\phi + \frac{2\sqrt{P_d}}{K} \sum_{n=jK}^{(j+1)K-1} d_n \Im_n \cos\phi n_Q(j) + n_Q^2(j) \right] \right\}$$

$$= \mathbf{E} \left\{ \left[P_c \cos^2\phi + \frac{2}{L} \sqrt{P_c} \cos\phi \sum_{j=0}^{L-1} n_I(j) + \frac{1}{L^2} \left(\sum_{j=0}^{L-1} n_I(j) \right)^2 \right] \right.$$

$$\left. \times \frac{1}{L} \sum_{j=0}^{L-1} \left[\frac{P_d}{K^2} \sum_{n=jK}^{(j+1)K-1} \sum_{m=jK}^{(j+1)K-1} d_n d_m \Im_n S_m \cos^2\phi + \frac{2\sqrt{P_d}}{K} \sum_{n=jK}^{(j+1)K-1} d_n \Im_n \cos\phi n_Q(j) + n_Q^2(j) \right] \right\}$$

$$= \frac{P_d P_c}{K} g_4 + P_c g_2 \frac{N_0}{2T_u} + \frac{P_d}{LK} g_2 \frac{N_0}{2T_u} + \frac{1}{L} \left(\frac{N_0}{2T_u}\right)^2 \tag{A-17}$$

JPL 1990-3: A 5-nrad Extragalactic Source Catalog Based on Combined Radio Interferometric Observations

O. J. Sovers
Tracking Systems and Applications Section

A combined analysis merges 17,000 DSN Very Long Baseline Interferometric (VLBI) observations with 303,000 observations from the Crustal Dynamics Project (CDP) and the International Radio Interferometric Surveying (IRIS) project. Observations from the Radio Reference Frame Development (RRFD) and Time and Earth Motion Precision Observations (TEMPO) programs through late 1990 form the DSN VLBI data set. The combined analysis yields angular coordinates of extragalactic radio sources with a precision of a few nanoradians, as compared with 5- to 10-nrad precision for coordinates derived in the past solely from DSN data. The improvement in the combined analysis is due to the new Mark III DSN data, as well as to increased statistical strength from the large volume of observations from non-DSN experiments. Such a unified analysis is made possible by recent improvements in parameter estimation software efficiency. The terrestrial reference frame is based on joint VLBI experiments employing both DSN and CDP antennas, and on specifying the coordinates of VLBI antennas in a proper geocentric coordinate system by means of Global Positioning System (GPS) collocation of VLBI, LLR, and SLR (Lunar and Satellite Laser Ranging) sites. Approximately 200 sources, fairly uniformly distributed over the sky at declinations ranging from −45 to +90 deg, form the 1990-3 catalog, and show formal uncertainties smaller than 5 nrad. At this level of precision, it is critical to estimate corrections to the International Astronomical Union (IAU) models of precession and nutation, and the 1990-3 source positions should only be used in conjunction with these corrections. Furthermore, at this level, a number of ignored effects, such as tropospheric turbulence and source structure, may cause errors equal to or greater than the formal uncertainties. Attempts to assess the influence of unmodeled systematic errors by comparing source positions with catalogs derived from independent analyses of partially independent data place a lower limit of 2 to 3 nrad on currently achievable accuracies. This leads to a realistic estimate of the true level of accuracy of approximately 5 nrad for this combined radio source catalog.

I. Introduction

This work emerged from a comparison of the models and analysis techniques employed by Goddard Space Flight Center (GSFC) and JPL groups [1]. Subsequently, it became of interest to determine whether adding non-DSN data to the analyses for radio-reference frame development could compensate for the reduced availability of the DSN antennas for Very Long Baseline Interferometry (VLBI) in the late 1980s. A further goal was to determine how much is to be gained by adding non-DSN data to the databases determining previously published JPL source coordinates [2,3], as well as to the more recent DSN-only data including several thousand Mark III observations [4]. By mid-1990 the total number of VLBI observations available for processing at JPL had exceeded 300,000. Recent improvements in JPL analysis software have made it possible to perform least-squares estimation involving hundreds of thousands of observations and tens of thousands of parameters within a few days of CPU time on VAX-780–class computers. It was therefore decided to generate a radio source catalog from all the VLBI data available in order to assess its suitability for meeting navigation requirements.

Three current projects demand celestial or terrestrial reference frames at or exceeding the accuracy levels of current frames. Proposed Galileo tracking requirements specify a radio reference frame stable to 5 nrad, while preliminary specifications for Ka-band tracking of Cassini tighten this requirement to 3 nrad. The TOPEX-POSEIDON project requires 3-cm three-dimensional baseline accuracy, with 5-cm alignment with the geocenter. This translates to 3 nrad on the California–Australia baseline. Thus, navigation requirements in the near future are more stringent than the accuracy level of 10 nrad for the DSN reference frame published in 1988 [2], and also go beyond the more recent DSN source catalogs [3,4].

The Crustal Dynamics Project (CDP) and International Radio Interferometric Surveying (IRIS) programs always perform multiple baseline experiments, with much more frequent observing sessions than those of the DSN. Consequently, the number of CDP+IRIS observations is more than an order of magnitude larger than the number of DSN observations generated by the Radio Reference Frame Development (RRFD) and Time and Earth Motion Precision Observations (TEMPO) projects at JPL. The CDP and IRIS networks also provide a considerably different geometrical coverage of the Earth than the three DSN complexes, and may thus serve to expose any systematic effects which depend on network geometry.

The combined VLBI database consists of 319,734 observations. In addition to being one of the largest data sets used for estimation of astrometric and geodetic parameters, it also provides a diverse set of baseline lengths and orientations. A long-standing weakness of all current VLBI data is the paucity of observations by stations in the southern hemisphere. This is especially true of the CDP and IRIS measurements throughout the 1980s; the DSN has benefitted from the Tidbinbilla station, which permits sky coverage down to −45-deg declination. On the other hand, all the non-DSN data have been acquired with Mark III recording systems, which have only recently (1988) been introduced into the DSN. The larger DSN antenna diameters and lower system temperatures, however, produce markedly reduced system noise and observable errors in current DSN Mark III data relative to CDP and IRIS data.

Aside from possible systematic errors, the radio source positions reported here form one of the best-determined extragalactic source catalogs presently available, which is a candidate for a fundamental reference frame. Discussions are under way in various International Astronomical Union (IAU) working groups, which are expected to adopt (in 1994) a radio reference frame to replace the present optical-based frame as the fundamental reference frame for astrometry. Such considerations are driven by the realization that uncertainties of the best optical measurements presently exceed the best radio interferometric uncertainties by at least an order of magnitude: <5 nrad (radio) versus ≈ 100 nrad (optical) [5]. When the next generation of optical astrometric measurements of galactic sources becomes available from the Hipparcos project [6], optical uncertainties are expected to be reduced to 10 nrad. Unlike the extragalactic radio sources, however, most galactic optical objects show considerable proper motion. As a result, the source positions reported here form a more stable reference frame than Hipparcos. During the next few years, additional experiments by the astrometry groups at the Naval Research Laboratory (NRL) and the U.S. Naval Observatory (USNO), as well as other current projects, are expected to improve precision further and to provide more nearly uniform sky coverage by including southern hemisphere observations.

This article presents a description of the analysis of combined VLBI observations. Section II summarizes the interferometric data from the four projects that are employed to generate the combined source catalog. Details such as distribution of observing sites, noise characteristics, and elevation angles serve to characterize the observations. VLBI modeling, analysis, and parametrization are discussed in Section III, while Section IV gives a sum-

mary of the results most pertinent to specification of radio source coordinates for spacecraft navigation. Section V attempts an assessment of the influence of unmodeled systematic errors in order to arrive at realistic estimates of the true level of accuracy of the combined radio source catalog.

II. VLBI Data

Only data from experiments employing hydrogen maser frequency standards and dual-frequency (S- and X-band, 2.3 and 8.4 GHz) ionosphere calibration are considered here. This means that no data prior to 1978 are included. Mark III VLBI systems [7] were used in acquiring all the CDP and IRIS data, and Block I systems [8] were used in TEMPO, while Mark II systems[1,2] were employed in RRFD experiments until late 1988 to early 1989, with Mark III used thereafter.

A short summary of the four observing programs that are considered in this article is presented in Table 1. Further details concerning data acquisition, reduction, and analysis are given in [9] for the GSFC CDP data, and in any current IRIS monthly bulletin [10] for the National Geodetic Survey IRIS data. Here, N_{stn} is the number of unique stations, and N_{bsl} is the number of unique baselines. Note that no distinction is made among various antennas at the three DSN complexes (marked with asterisks). The number of observing sessions is N_{ses}; these are normally of 24-hr duration (except for the post-1987 RRFD and the 3-hr TEMPO sessions), and may involve as many as seven stations. The number of unique sources observed in each program is N_{src}. The major thrust of the first three observing programs of Table 1 is geodetic, while the RRFD program is predominantly astrometric, as can be seen from the N_{bsl} and N_{src} columns. An order of magnitude more observations of relatively few sources by CDP, and especially IRIS, may contribute more to determination of short-period nutation amplitudes than to improving the accuracy of source positions. Finally, N_{obs} is the number of pairs of delay (D) and delay-rate (DR) measurements that are included in the analysis. The analysis described here used data available in the fall of 1990; by early spring of 1991, the number of RRFD MkIII observations had tripled, and they are being incorporated into the database for a future source catalog. A total of 230 extragalactic radio sources contributed observations. The number of observations per source ranges from a maximum of 23,656 (for 4C 39.25) to 1 (for five sources for which RRFD observations had only recently begun).

Two broad characterizations of the data involve data-noise and elevation-angle distributions. Table 2 shows the root-mean-square delay data noise (formal uncertainty from correlation processing) for each observing program, as well as the additional noise (constant for each baseline in a given observing session) that was root-sum-squared with the observable noise in order to approximate chi-square per degree of freedom = 1 in the least-squares fit (see Section III). Note that for TEMPO data, additional noise is introduced as an RSSed fraction of the total tropospheric delay contribution for each observation [4], rather than as an RSSed constant for each observing session. The RRFD data are subdivided into Mark II and Mark III categories because there is a large difference in their data quality. While the data quality also generally improves with time from the late 1970s to 1990 for all observing programs, there is no such clear demarcation for CDP, IRIS, or TEMPO.

It can be seen that both the data noise and additive noise are considerably lower for CDP and IRIS data than for RRFD MkII and TEMPO data. The RRFD MkIII data, on the other hand, exhibit the lowest value in both categories. Comparison of data noise and additive noise is not straightforward because of variations in scan lengths and fractions of low-elevation observations among the observing programs. Variation of the additive noise may indicate problems either with the receiving or recording systems, correlation processing, data editing, or the post-correlation software that generates the observables. Noise added to account solely for unmodeled physical effects is not expected to vary with the type of data if similar observing schedules are used. The strength of the DSN data as a whole lies in the larger antennas and variety of sources observed, as well as the generally longer baselines and more complete declination coverage. The DSN observations also benefit from quieter receiving electronics, given equivalent recording systems. These characteristics permit the 17,000 DSN observations to make a nearly equivalent contribution, on the whole, to that of the 300,000 CDP+IRIS observations in the combined fit.

Low-elevation-angle observations were avoided in early non-DSN VLBI experiments because of problems in modeling the large tropospheric delays at low elevations. As a result, there is a scarcity of data at elevation angles below 15–20 deg in the early CDP and IRIS data. The DSN experiments, partly because they were performed over longer baselines, routinely involved observations down to the an-

[1] E. J. Cohen, "VLBI Bandwidth Synthesis Manual," JPL Tracking Systems and Applications Section report (internal document), June 1979.

[2] E. J. Cohen, "VLBI Rack Manual," JPL Tracking Systems and Applications Section report (internal document), May 1980.

tenna limits of 6-deg elevation throughout their entire history. Now that improved troposphere models are available [11,12], and it is commonly accepted that mismodeling of the time-varying tropospheric delay is a limiting error source in interpretation of VLBI results [13], and that low-elevation observations help to model and reduce the static part of this error, CDP and IRIS experiments also routinely acquire substantial numbers of low-elevation observations. Table 3 shows an overall summary of the four observing programs in terms of the fractions of observations involving elevation angles smaller than 10, 20, and 30 deg (f_{10}, f_{20}, and f_{30}, respectively) at any participating station. It may be seen that both f_{10} and f_{20} are larger by approximately a factor of 3 for RRFD and TEMPO than for CDP and IRIS observations. Well over half of the DSN observations are made at elevations below 30 deg.

Two analyses of comparably large VLBI data sets have been recently carried out by Ma and co-workers [14,15]. These two catalogs are based respectively on 237,000 and 461,000 observations, with the second including NAVNET (the geodetic VLBI program of the USNO) and southern hemisphere astrometric data in addition to CDP and IRIS experiments. There is necessarily considerable overlap between those two databases and the one used in the work described here. Comparison of estimated source positions indicates the magnitude of discrepancies due to differences in observables, modeling software, and analysis techniques.

III. Modeling and Data Analysis

With some exceptions, theoretical modeling of the observed delays and rates conforms to the specifications contained in the International Earth Rotation Service (IERS) standards [16]. The data quality demands a model for the Earth's precession and nutation that goes beyond the current IERS and IAU standards [17]. Details are given below, and a more complete description of modeling is given in the document concerning the JPL VLBI software "MODEST" [18]. Briefly, the calculations are performed in solar system barycentric coordinates defined in terms of the mean equator of J2000.0.

The clock and troposphere modeling options were the same as those used in the 1988 source catalog paper [2]. The clock model was piecewise linear, with breaks introduced only in places where they appeared essential from inspection of the post-fit residual patterns. At most stations, this averaged to 5 or 6 clock sections per 24 hr. Improved instrumental stability reduces this by at least a factor of 2 for RRFD Mark III data. Troposphere modeling was piecewise constant, with a new value of the zenith delay estimated at each station for every 3 hr of observations. It remains to be determined how much error in source coordinates is introduced by such parametrization.

UT1 and polar motion (UTPM) values were fixed at those given by the uniformly smoothed time series of Gross and Steppe [19], which is consistent with the 1989 IERS UTPM reference frame [20]. These Earth-rotation parameters are accurate to 1 to 2 nrad during the latter part of the time span covered by the VLBI data. The coordinates of the reference station (Gilmore Creek, Alaska) were taken from the CDP Annual Report for 1989 [9] at the epoch 1988.0. These coordinates are specified in terms of a current best-estimate geocentric system, established in recent years by means of comparison of VLBI, SLR, LLR, and GPS measurements [21]. Positions at epoch 1988.0 are estimated for 30 stations (all but the reference station). Except for Gilmore Creek, time rates of change are also estimated for the coordinates of all stations for which the data cover a sufficient time span (19 stations in all).

Note that the adoption of a terrestrial reference frame poses a consistency problem in analysis of large data sets. Considerable portions of the data used here have also gone into the results of all the references cited in the previous paragraph. Achieving truly independent fixes of the terrestrial coordinate system and the UTPM series would necessitate the adoption of results of techniques totally independent of VLBI. Satellite and lunar ranging techniques, however, are presently not capable of providing all the required parameters at an acceptable accuracy level.

Similar considerations apply, to a somewhat lesser degree, to the method used to circumvent the right ascension reference problem for VLBI data. The origin of celestial coordinates was fixed by tightly constraining the right ascension and declination of OJ 287 and declination of CTD 20 to their values in the IERS 90C01 celestial reference frame [20]. Instead of the usual single right ascension constraint, three coordinates are constrained in order to allow for improved nutation and precession models, and to make the overall orientation of the estimated coordinate system be close to IERS 90C01. It was noted from previous experience that alternative choices of terrestrial and celestial origins have little effect on the relative source coordinates. Nutation modeling is a known weak point of the present IERS standards, which use the 1980 IAU nutation series [17]. The position of the celestial pole is known to depart from the 1980 IAU value by as much as 25 nrad [4,22,23]. The quality of present VLBI data demands that this inadequacy be compensated by some means. Consequently, it was decided to estimate values of the precession constant and selected nutation amplitudes.

Ecliptic longitude (ψ) and obliquity (ε) define the position of the celestial ephemeris pole (CEP) at a given epoch. The estimated components of the nutation model were the amplitudes of the nutations with selected periods in both longitude and obliquity. There is evidence [24] that effects of ocean tides and friction at the core–mantle boundary cause small contributions to nutation that are out of phase with those of the current IAU model, which was derived for an oceanless Earth. Consequently both in-phase and out-of-phase components were considered. Amplitude corrections ΔA_ψ and ΔA_ε of a total of nine nutation periods were estimated: 18.6 yr, 9.3 yr, 365.3 d, 182.6 d, 121.7 d, 27.6 d, 13.7 d, 13.6 d (respectively, terms 1, 2, 10, 9, 11, 32, 31, and 33 of the total of 106 in the 1980 IAU series), and the free core nutation with a period of 433.2 d. These particular frequencies include at least all the terms which have been found to require amplitude corrections in previous work [22,23]. Exploratory fits on various subsets of the present database have shown that the 18-yr, annual, and semiannual terms give a good representation of the daily corrections to the CEP position ($\Delta\psi$ and $\Delta\varepsilon$) determined from the data. The remaining terms are included here in order to uncover other, possibly significant corrections. With 4 components per period, the total number of estimated nutation amplitudes was 36. In order to constrain the celestial ephemeris pole at epoch J2000 to its nominal position, two angular offsets (rotations about the x and y axes) were also estimated. These represent, respectively, displacements of the CEP toward 18- and 0-hr right ascension. The method used here for defining the celestial reference frame and for estimating precession and nutation follows [4].

A linear least-squares algorithm is used to obtain optimal estimates of the model parameters from data. The least-squares estimation is achieved with the square-root information filter (SRIF) algorithms of Bierman [25]. The software is presently implemented on VAX computers and workstations, which were used in the fit described in this article. Several days to a week of CPU time is required to complete the analysis. Disk storage requirements are also quite sizable ($\approx$0.5 Gbytes), mostly due to the large number of source coordinates that enter each fit as global parameters. To minimize CPU and storage requirements, the analysis was done in six sections: two CDP, three IRIS, and one RRFD/TEMPO, with no section containing more than 70,000 observations. The basic fit estimates 57,795 parameters from the 319,734 pairs of delay and delay-rate observables. Of these, only the 460 source coordinates, 39 nutation/precession quantities, and 294 station positions and rates were global parameters. The final combined results are identical to what would be obtained by performing the fit in a single step [25].

The least-squares parameter estimation is weighted by a covariance matrix of the observables that is diagonal. The diagonal elements are inversely proportional to the square of the observable error. Such effects as uncalibrated instrumental errors, troposphere fluctuations [13], and source structure [26–28] introduce additional unmodeled noise. In order to partially account for these inadequacies, session-specific adjustable errors are introduced. As mentioned in Section II, these are root-sum-squared with the observable errors, and adjusted in order to make the normalized chi-square for each session close to 1.0. For delays, this additive noise ranges from a few to tens of picoseconds for Mark III observations, to several hundred picoseconds for some of the early Mark II and Block I sessions. As mentioned before, the variation of additive noise may indicate possible problems in instrumentation, correlation, or postcorrelation processing that need to be investigated further.

IV. Results

The overall goodness of fit is indicated by root-mean-square (RMS) residuals of 123 psec for delays, and 100 fsec/sec for delay rates. These are distributed among the component data sets, as shown in Table 4. Generally, the residuals fall into two classes: Mark II (TEMPO and RRFD) and Mark III (CDP, IRIS, and RRFD). The noticeably better IRIS DR residuals may originate in the smaller proportion of IRIS low-elevation observations. Since this article focuses on the celestial reference frame, three categories of estimated parameters (clock, troposphere, and station location) will not be considered in detail here. The first two categories (clocks and tropospheres) will be subject to future scrutiny in order to better characterize their stochastic behavior. Station coordinates at epoch 1988.0 are generally determined with formal uncertainties of $\approx$2 to 20 mm, and their linear time rates of change with uncertainties of $\approx$1 to 5 mm/yr. The rates agree with the Minster-Jordan AM0-2 [29] and NUVEL [30] models of tectonic motion at the level of a few mm/yr. Naturally there is an intimate relationship between the models of UTPM and tectonic motion, which needs to be explored more critically in future work.

Not counting the two reference sources (OJ 287 and CTD 20), coordinate estimates are made for a total of 230 sources. Positions of 216 of these sources are listed in Table 5, including the number of observations, average observing epoch, and correlation coefficient between right ascension and declination. Note that the source coordinates are given in the traditional units of hours, minutes, and seconds for RA, and degrees, minutes, and seconds for dec-

lination (δ). Approximate conversion factors to nanoradians are $72\cos(\delta)$ nrad/msec for RA and 4.8 nrad/mas for δ. Fourteen sources with 1σ declination uncertainties that exceed $\approx$20 nrad are not included in Table 5. On the other hand, the Table includes 23 sources with fewer than 10 observations, whose coordinates should be considered preliminary. Most of these are sources for which DSN observations have been initiated only recently. Nevertheless, approximately half of these sources have σ's smaller than 5 nrad. To give the reader some idea of the quality of this catalog, it contains 40 sources that were observed more than 1000 times, 54 sources with formal declination uncertainties (σ_δ) smaller than 1 nrad, and a majority (178) of the sources have $\sigma_\delta < 5$ nrad. The catalog is named JPL 1990-3, in conformance with the local naming convention which categorizes catalogs by date of analysis.

Figures 1(a) and (b) show histograms of the formal uncertainties of JPL 1990-3 divided into 0.25-nrad bins. For comparison, similar histograms are shown in Figs. 2(a) and (b) for the current TEMPO catalog, JPL 1991-1 [4], which is based on DSN data through early 1991. It is seen that the errors of 1991-1 bear some resemblance to single-peaked distributions, while the heterogeneous nature of the observations entering 1990-3 spreads out the error distributions considerably and introduces multiple peaks.

Table 6 shows the precession and nutation parameters (in units of nanoradians) derived from the fit which yield JPL 1990-3. Note the highly significant precession correction. The x and y celestial ephemeris pole offsets are listed in the ΔA_ψ and ΔA_ε columns, and show that the CEP at J2000 is significantly shifted, approximately toward 5.5 hours RA. There are sizable significant corrections to the 1980 IAU nutation amplitudes at 18-yr, 9-yr, annual, and semiannual periods, including a number of out-of-phase components. A number of the shorter-period nutation amplitudes are not significant, judging by the given formal uncertainties, but are included for completeness. A long-standing problem in estimating precession and nutation quantities is the short timebase of VLBI observations relative to the 18.6-yr nutation period. This leads to high correlations between precession and nutation, and the problem is not expected to be fully removed until the data cover a full 18.6-yr cycle in the late 1990s. The formal uncertainties in Table 6 should be interpreted with these high correlations in mind (the largest correlation coefficient between precession and nutation is 0.93). Nevertheless, there is substantial agreement of these values with comparable theoretical [31,32] and experimental [22,33] estimates. Note that the quoted experimental results use data that partly overlap the data used here. It is stressed that any application using the source coordinates of Table 5 must also model the celestial pole position with the parameters of Table 6.

V. Consistency Tests and Accuracy Estimates

The database involved in generating the radio source coordinates presented here is too large to permit many internal consistency tests to be carried out within a reasonable time span, with the computing facilities currently available. Therefore, assessment of the validity of the formal 1σ statistical parameter uncertainties was performed by comparison of the estimated coordinates with those of a variety of other catalogs, based on data that overlap the present database to varying degrees.

The first category of such catalogs includes those based only on RRFD and TEMPO data. The DSN catalogs JPL 1987-1 [2] and JPL 1991-1 [4] are selected for comparison with JPL 1990-3. Catalog JPL 1987-1 contains positions of 128 sources, and is the last DSN catalog described in an external refereed publication. Note that data entering 1987-1 terminate in late 1985, well before the inception of DSN Mark III systems. Catalog JPL 1991-1 is the current TEMPO catalog, based on RRFD and TEMPO data through early 1991, and thus includes more sources (241) and more than twice the number of RRFD Mark III observations used for 1990-3.

The second category of catalogs includes those determined by other research groups. There are many possible candidates for comparison, of which only three are chosen. One important member of this category is the IERS catalog 90C01 [20], which combines individual CDP, IRIS, USNO, and JPL catalogs, with due attention paid to removal of rotational offsets and evaluation of formal uncertainties from disparate analyses. It contains the coordinates of 228 sources. Two GSFC catalogs were chosen to be representative of other external results. The first, here denoted GSFC89, is from [14]. It is based on 238,000 CDP and IRIS observations through early 1988 and contains 182 sources. A somewhat premature attempt was made to tie these source coordinates to the optical FK5 frame, which accounts for the large ($\approx$25 nrad) rotation in right ascension that is required to make this catalog coincide with any conventional radio catalog. A later Goddard catalog, denoted GSFC90 [15], extends the CDP and IRIS database and adds NAVNET [34] and southern hemisphere astrometric experiments, for a total of 461,000 observations of 318 sources.

Pairwise comparisons were made between the present source catalog and the above five collections of source

positions derived from VLBI data. Prior to comparing the source coordinates, three-dimensional rotational offsets were calculated and removed for each pair in order to account for the different right ascension and nutation reference points used in deriving each catalog. These rotational offsets are a manifestation of the different methods used to define the orientation of the celestial reference frame for each catalog. Tables 7 and 8 give the magnitudes of these rotational offsets (whose uncertainties are of the order of 0.1 nrad), as well as the χ^2 per degree of freedom (χ^2_ν) in the least-squares solution determining the rotational transformations. After the rotation is applied (to the catalog other than 1990-3), the right ascension differences are scaled to arc-length differences by applying the factor cos (declination), and then RMS differences are calculated for both source coordinates. Note that all the catalogs involved in the comparisons (with the exception of 1987-1) were generated by fits that corrected the 1980 IAU nutation and precession errors by estimating either daily CEP orientation or nutation amplitudes and the precession constant.

Statistical analysis of uncertainties is problematic in cases like the present, where the same parameters are estimated from partially overlapping data. When this work is extended to include databases through 1991, a more complete statistical assessment will need to be performed. For the present, the covariance matrices for both catalogs are assumed to be diagonal, and the uncertainty of the difference of a given coordinate is assumed to be the root sum square of the individual 1σ uncertainties. In previous work, neglecting off-diagonal covariances somewhat underestimated the true errors.

From Table 7, it can be seen that agreement of JPL 1990-3 with the other two JPL catalogs is reasonable, at the level of the formal uncertainties. Rotational offsets are as large as 7 nrad for 1987-1, but much smaller for the more recent 1991-1. This is a manifestation of precession/nutation model deficiencies that were not as well corrected for 1987-1 as in more recent fits. Root-mean-square RA and declination differences are nearly consistent with the uncertainty level of the catalogs ($\approx$10 nrad for 1987-1 and $<$ 5 nrad for 1991-1 and 1990-3). The χ^2_ν(RA) values of 2.0 and 1.8 indicate some problem with right ascensions that is not encountered with declinations. In order to test whether the comparisons are being distorted by selection of sources, the two comparisons were repeated, including only sources that have formal σ's smaller than 5 nrad in both catalogs. Results for rotational offsets and χ^2_ν's are essentially unchanged, and the RMS differences decrease to approximately 4 and 3 nrad for 1987-1 and 1991-1, respectively.

Comparisons with the IERS catalog and two GSFC catalogs in Table 8 paint a more pessimistic picture than the intra-JPL comparisons. Rotational offsets are as large as 10 nrad (the 24-nrad rotation of GSFC89 was already mentioned above), RMS differences are slightly higher than 5 nrad, and χ^2_ν's range from 2 to 4. A 5-nrad cut on input catalog errors decreases the RMS differences slightly but raises the χ^2_ν's even further. This suggests that the formal uncertainties are underestimating true errors due to some systematic effects coming into play at the nanoradian level. When the catalog formal σ's are root sum squared with 2 nrad for all coordinates of each catalog pair in the comparisons, the χ^2_ν's are reduced to the 0.8–1.4 range; a slightly larger additive error (3 nrad) is required in the IERS90 comparison. Until further work identifies the source of the 2–3-nrad systematic errors, these results place a lower limit on the accuracy achievable in source position determinations. To be investigated are tropospheric covariance, the consequences of UTPM errors; a detailed comparison of modeling and analysis procedures, and proper motion or time-varying source structure. Approximately 100 sources show formal position uncertainties of 2 nrad or smaller in the JPL 1990-3 catalog. Indications are that the low σ's are solely due to the extremely large number of observations, and do not properly reflect presently unidentified and unmodeled systematic errors.

In conclusion, the addition of several hundred thousand non-DSN VLBI observations to the DSN data set in a fit determining a radio reference frame has the beneficial effects of checking consistency of source coordinates derived from disparate sets of baselines and revealing modeling deficiencies at the 2-nrad level. Considerable improvement is seen in the formal uncertainties (Fig. 1 versus Fig. 2), nearly 50 percent of the 1990-3 uncertainties are brought below the level of presently unidentified systematic effects. A similar level of resolution may not have been reached solely with DSN data for another year or two. It appears that the true accuracy of the JPL 1990-3 catalog is approximately at the 5-nrad level.

VI. Conclusions

Source angular coordinates of 216 extragalactic objects are derived from a database that includes the majority of the world's astrometric VLBI measurements through 1990. Care has been taken to eliminate some known sources of systematic error, in particular the inadequacy of the present standard precession/nutation model. The formal uncertainties of these JPL 1990-3 source positions are predominantly smaller than 5 nrad. Consistency tests indicate that the formal errors are overoptimistic, espe-

cially those below 2 to 3 nrad. The 2- to 3-nrad limit appears to be the point at which unknown systematics dominate. In spite of these systematics, comparisons with source coordinates from independent analyses of data that partly overlap the database used here show that the 1990-3 positions agree at approximately the 5-nrad level. This catalog may thus be one of the first to exhibit true 5-nrad accuracy. Consistency with the non-DSN data in the analysis, with its greater variety of baseline orientations and lengths, indicates that no serious systematic errors are present in the DSN-only data that might be introduced by the DSN baseline geometry, particularly the extremely long California–Australia baseline. Consistency with GSFC catalogs also provides an indirect verification that there are no outstanding discrepancies between JPL and GSFC modeling and estimation software at the 5-nrad level.

The status of the 1990-3 reference frame relative to future mission requirements remains to be fully characterized. When used with the CEP model of Table 6, the 1990-3 source coordinates are expected to yield accurate source coordinates of date several years into the future. If the true accuracy of this catalog is at the 5-nrad level, as indicated by the above-mentioned comparisons, then it is very close to satisfying Galileo specifications. The TOPEX/POSEIDON S-/X-band and preliminary Cassini Ka-band (32-GHz) 3-nrad requirements are more difficult to meet; however, many of the errors in constructing a Ka-band reference frame should be the same as those for the present S-/X-band frame. It is expected that further analytical refinements and acquisition of additional VLBI measurements will permit achieving the 3-nrad level within the next few years.

Future data analysis is planned along lines similar to that presented here, incorporating the more recent high-quality DSN Mark III data, as well as updates of the TEMPO, CDP, and IRIS databases. Empirical corrections for two of the known causes of systematic error (source structure and troposphere mismodeling) will be incorporated in future parameter estimates. Including LLR data extending back to 1969 would also be of substantial help in reducing the precession-nutation correlations [33]. Such fits are expected to reduce source coordinate uncertainties further, to better characterize the true accuracies, and to satisfy future navigation requirements for a radio reference frame.

Acknowledgments

Members of the Astrometric Techniques Group performed the data collection and correlation analysis of the RRFD experiments. In particular, S. T. Lowe, C. S. Jacobs, R. N. Treuhaft, R. F. Coker, and D. N. Fort participated in the difficult transition from Mark II to Mark III processing. R. P. Branson implemented a restart capability in the parameter estimation code. J. A. Steppe and S. H. Oliveau contributed the TEMPO data, C. Ma of GSFC arranged for R. Kennard of the CDP Data Information System to supply the CDP intercontinental measurements, and M. Morrison of NGS supplied the IRIS data. During the summer of 1987, C. K. Dang, J. H. Lieske, Jr., and M. C. Meyer helped in the initial stages of the data analysis. C. S. Jacobs and R. N. Treuhaft made many useful comments concerning the manuscript.

References

[1] O. J. Sovers and C. Ma, "Comparison of GSFC and JPL VLBI Modeling Software: Benchmark," *TDA Progress Report 42-83*, vol. July–September 1985, Jet Propulsion Laboratory, Pasadena, California, pp. 101–112, November 15, 1985.

[2] O. J. Sovers, C. D. Edwards, C. S. Jacobs, G. E. Lanyi, K. M. Liewer, and R. N. Treuhaft, "Astrometric Results of 1978–1985 Deep Space Network Radio Interferometry: The JPL 1987-1 Extragalactic Source Catalog," *Astron. J.*, vol. 95, no. 6, pp. 1647–1658, 1988.

[3] J. S. Ulvestad, O. J. Sovers, and C. S. Jacobs, "A Higher-Density VLBI Catalog for Navigating Magellan and Galileo," *TDA Progress Report 42-100*, vol. October–December 1989, Jet Propulsion Laboratory, Pasadena, California, pp. 274–300, February 15, 1990.

[4] J. A. Steppe, S. H. Oliveau, and O. J. Sovers, "Earth Rotation Parameters from DSN VLBI: 1991," submitted to *International Earth Rotation Service, Annual Report for 1990*, Paris, France, 1991.

[5] H. Schwan, "The FK5: Present Status and Some Derived Results," *IAU Symposium 141 Proceedings (Kluwer)*, Leningrad, USSR, 1989.

[6] M. A. C. Perryman, "Astrometry from Space," *Nature*, vol. 340, no. 6229, pp. 111–116, July 13, 1989.

[7] A. E. E. Rogers, R. J. Cappallo, H. F. Hinteregger, J. I. Levine, E. F. Nesman, J. C. Webber, A. R. Whitney, T. A. Clark, C. Ma, J. Ryan, B. E. Corey, C. C. Counselman, T. A. Herring, I. I. Shapiro, C. A. Knight, D. B. Shaffer, N. R. Vandenberg, R. Lacasse, R. Mauzy, B. Rayhrer, B. R. Schupler, and J. C. Pigg, "Very-Long-Baseline Interferometry: The Mark III System for Geodesy, Astrometry, and Aperture Synthesis," *Science*, vol. 219, no. 4580, pp. 51–54, January 7, 1983.

[8] J. V. Lu Valle, R. D. Shaffer, M. G. Roth, T. M. Eubanks, and P. S. Callahan, "Operational VLBI Clock Synchronization and Platform Parameter Determination," *TDA Progress Report 42-66*, vol. September–October 1981, Jet Propulsion Laboratory, Pasadena, California, pp. 307–318, December 15, 1981.

[9] D. S. Caprette, C. Ma, and J. W. Ryan, "Crustal Dynamics Project Data Analysis—1990," NASA TM 100765, NASA, Washington, D.C., December 1990.

[10] International Association of Geodesy, *International Radio Interferometric Surveying (IRIS) Bulletin A*, no. 88, June 1991.

[11] G. E. Lanyi, "Tropospheric Delay Effects in Radio Interferometry," *TDA Progress Report 42-78*, vol. April–June 1984, Jet Propulsion Laboratory, Pasadena, California, pp. 152–159, August 15, 1984.

[12] J. L. Davis, T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered, "Geodesy by Radio Interferometry: Effects of Atmospheric Modeling Errors on Estimated Baseline Length," *Radio Science*, vol. 20, no. 6, pp. 1593–1607, November–December 1985.

[13] R. N. Treuhaft and G. E. Lanyi, "The Effect of the Dynamic Wet Troposphere on Radio Interferometric Measurements," *Radio Science*, vol. 22, no. 2, pp. 251–265, March–April 1987.

[14] C. Ma, D. B. Shaffer, C. DeVegt, K. J. Johnston, and J. L. Russell, "A Radio Optical Reference Frame, I. Precise Radio Source Positions Determined by

VLBI: Observations from 1979 to 1988 and a Tie to the FK5," *Astron. J.*, vol. 99, no. 4, pp. 1284–1298, April 1990.

[15] C. Ma and D. B. Shaffer, "Stability of the Extragalactic Reference Frame Realized by VLBI," *Proceedings of the 127th Colloquium of the IAU: Reference Systems*, Virginia Beach, Virginia, 1990, pp. 135–144, U. S. Naval Observatory, 1991.

[16] D. D. McCarthy, ed., "IERS Standards, IERS Technical Note 3," *International Earth Rotation Service*, pp. 37–44, Paris, France, 1989.

[17] P. K. Seidelmann, "1980 IAU Theory of Nutation: The Final Report of the IAU Working Group on Nutation," *Celestial Mechanics*, vol. 27, no. 1, pp. 79–106, 1982.

[18] O. J. Sovers, "Observation Model and Parameter Partials for the JPL VLBI Parameter Estimation Software MODEST—1991," JPL Publication 83-39, Rev. 4, Jet Propulsion Laboratory, Pasadena, California, August 15, 1991.

[19] R. S. Gross and J. A. Steppe, "A Combination of Earth Orientation Data: SPACE90," submitted to *International Earth Rotation Service, Annual Report for 1990*, Paris, France, 1991.

[20] *International Earth Rotation Service, Annual Report for 1989*, Paris, France, 1990.

[21] J. R. Ray, C. Ma, J. W. Ryan, T. A. Clark, R. J. Eanes, M. M. Watkins, B. E. Schutz, and B. D. Tapley, "Comparison of VLBI and SLR Geocentric Site Coordinates," *Geophys. Res. Lett.*, vol. 18, no. 2, pp. 231–234, February 1991.

[22] T. A. Herring, C. R. Gwinn, and I. I. Shapiro, "Geodesy by Radio Interferometry: Studies of the Forced Nutations of the Earth, Part 1: Data Analysis," *J. Geophys. Res.*, vol. 91, no. B5, pp. 4745–4754, April 10, 1986.

[23] S. Y. Zhu, E. Groten, and C. Reigber, "Various Aspects of Numerical Determination of Nutation Constants, II. An Improved Nutation Series for the Deformable Earth," *Astron. J.*, vol. 99, no. 3, pp. 1024–1044, March 1990.

[24] C. R. Gwinn, T. A. Herring, and I. I. Shapiro, "Geodesy by Radio Interferometry: Studies of the Forced Nutations of the Earth, Part 2: Interpretation," *J. Geophys. Res.*, vol. 91, no. B5, pp. 4755–4765, April 10, 1986.

[25] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, New York: Academic Press, 1977.

[26] J. S. Ulvestad, "Effects of Source Structure on Astrometry and Geodesy," *IAU Symposium 129 Proceedings (Reidel)*, Cambridge, Massachusetts, 1988.

[27] P. Charlot, "Structure des Sources Radio Extragalactiques dans les Observations VLBI d'Astrometrie et de Geodynamique," PhD thesis, Paris Observatory, Paris, France, 1989.

[28] P. Charlot, "Radio Source Structure in Astrometric and Geodetic Very Long Baseline Interferometry," *Astron. J.*, vol. 99, no. 4, pp. 1309–1326, April 1990.

[29] J. B. Minster and T. H. Jordan, "Present-Day Plate Motions," *J. Geophys. Res.*, vol. 83, no. B11, pp. 5331–5354, 1978.

[30] C. DeMets, R. G. Gordon, D. F. Argus, and S. Stein, "Current Plate Motions," *Geophys. J. Int.* (in press), 1990.

[31] S. Y. Zhu and E. Groten, "Various Aspects of Numerical Determination of Nutation Constants, I. Improvement of Rigid-Earth Nutation," *Astron. J.*, vol. 98, no. 3, pp. 1104–1111, September 1989.

[32] H. Kinoshita and J. Souchay, "The Theory of the Nutation for the Rigid Earth Model at the Second Order," *Celestial Mech. and Dyn. Astron.*, vol. 48, no. 3, pp. 187–265, 1990.

[33] P. Charlot, O. J. Sovers, J. G. Williams, and X X Newhall, "A Global VLBI/LLR Analysis for the Determination of Precession and Nutation Constants," *Proceedings of the 127th Colloquium of the IAU: Reference Systems*, Virginia Beach, Virginia, 1990, pp. 228–233, U. S. Naval Observatory, 1991.

[34] T. M. Eubanks, M. S. Carter, F. J. Josties, D. N. Matsakis, and D. D. McCarthy, "The Radio Reference Frame of the U. S. Naval Observatory Radio Interferometry Program," *Proceedings of the 127th Colloquium of the IAU: Reference Systems*, Virginia Beach, Virginia, 1990, pp. 256–260, U. S. Naval Observatory, 1991.

Table 1. Summary of VLBI observing programs

Program	N_{stn}	N_{bsl}	Time period	N_{ses}	N_{src}	N_{obs}
CDP	31	144	79/ 8 – 89/12	178	76	134,627
IRIS	8	24	84/ 1 – 90/ 6	452	31	168,119
TEMPO	3*	3*	80/ 7 – 90/ 9	665	106	7,324
RRFD MkII	3*	2*	78/10 –89/ 4	65	230	8,520
RRFD MkIII	3*	2*	88/ 8 – 90/ 7	13	214	1,144
Total	34	147	78/10 – 90/ 9	1373	230	319,734

*DSN Deep Space Communications Complex

Table 2. Noise characteristics of VLBI data

Program	N_{obs}	RMS data noise, psec	RMS added noise, psec
CDP	134,627	100	77
IRIS	168,119	94	66
TEMPO	7,324	464	260
RRFD MkII	8,520	395	158
RRFD MkIII	1,144	65	50

Table 3. Elevation-angle distributions of VLBI data

Program	f_{10}	f_{20}	f_{30}
CDP	0.020	0.129	0.278
IRIS	0.023	0.111	0.218
TEMPO	0.066	0.310	0.548
RRFD	0.065	0.293	0.510

Table 4. RMS delay and rate residuals of VLBI data

Program	Delay, psec	Delay rate, fsec/sec
CDP	96	113
IRIS	86	81
TEMPO	448	109
RRFD MkII	336	134
RRFD MkIII	87	117
All	123	100

Table 5. Source coordinates from combined VLBI measurements, 1978–90

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
0008−264	P 0008−264	1984.667	38	0	11	1.246868	0.071	−26	12	33.37761	0.81	−0.933
0013−005	P 0013−00	1989.589	8	0	16	11.088443	0.151	−0	15	12.44239	2.25	−0.989
0016+731	0016+731	1988.839	33	0	19	45.785908	0.131	73	27	30.01964	0.39	0.102
0019+058	P 0019+058	1984.773	50	0	22	32.441219	0.026	6	8	4.26917	0.77	−0.570
0048−097	P 0048−09	1989.463	344	0	50	41.317356	0.007	−9	29	5.20976	0.12	−0.195
0104−408	P 0104−408	1985.701	98	1	6	45.108147	0.060	−40	34	19.95992	0.45	−0.875
0106+013	P 0106+01	1986.699	12430	1	8	38.771065	0.002	1	35	0.31787	0.06	−0.088
0111+021	P 0111+021	1983.353	37	1	13	43.145073	0.098	2	22	17.31556	1.58	−0.913
0112−017	P 0112−017	1988.929	24	1	15	17.099895	0.015	−1	27	4.57657	0.29	−0.779
0113−118	P 0113−118	1984.385	61	1	16	12.521992	0.035	−11	36	15.43412	0.52	−0.902
0119+115	P 0119+11	1988.888	22	1	21	41.595005	0.015	11	49	50.41420	0.32	−0.761
0119+041	GC 0119+04	1990.096	1023	1	21	56.861675	0.004	4	22	24.73470	0.15	−0.252
0133+476	DA 55	1984.855	172	1	36	58.594818	0.010	47	51	29.10011	0.25	−0.365
0146+056	0146+056	1988.710	14	1	49	22.370842	0.033	5	55	53.57026	0.61	−0.859
0149+218	P 0149+21	1989.085	8	1	52	18.059040	0.048	22	7	7.70014	0.77	−0.815
0201+113	P 0201+113	1987.110	15	2	3	46.657017	0.025	11	34	45.41072	0.56	−0.748
0202+149	P 0202+14	1985.729	123	2	4	50.413879	0.008	15	14	11.04385	0.15	−0.701
0208−512	P 0208−512	1988.393	3	2	10	46.199532	0.271	−51	1	1.88993	1.83	−0.886
0212+735	0212+735	1987.203	20017	2	17	30.813374	0.009	73	49	32.62226	0.03	−0.437
0221+067	GC 0221+06	1988.948	12	2	24	28.428144	0.027	6	59	23.34205	0.38	−0.900
0224+671	DW 0224+67	1981.836	868	2	28	50.051556	0.014	67	21	3.02982	0.10	−0.386
0229+131	P 0229+13	1987.214	11202	2	31	45.894047	0.002	13	22	54.71671	0.04	−0.164
0234+285	CTD 20	1987.195	6544	2	37	52.405674	0.003	(28	48	8.99038)	(0.00)	(0.000)
0235+164	GC 0235+16	1982.164	903	2	38	38.930103	0.006	16	36	59.27496	0.11	−0.666
0239+108	OD 166	1985.359	76	2	42	29.170900	0.030	11	1	0.72751	0.37	−0.931
0250+178	GC 0250+17	1989.677	12	2	53	34.882183	0.037	18	5	42.52564	1.05	−0.566
0256+075	OD 094.7	1983.652	32	2	59	27.076651	0.036	7	47	39.64091	1.83	−0.399
0300+470	OE 400	1988.855	3466	3	3	35.242237	0.004	47	16	16.27578	0.04	−0.306
0306+102	0306+102	1989.282	14	3	9	3.623471	0.028	10	29	16.34191	0.42	−0.895
0309+411	0309+411	1987.721	12	3	13	1.962166	0.035	41	20	1.18315	1.01	−0.157
0316+413	3C 84	1984.601	308	3	19	48.160145	0.010	41	30	42.10449	0.18	−0.044
0317+188	P 0317+188	1989.671	11	3	19	51.256706	0.012	19	1	31.29180	0.32	−0.646
0326+277	0326+277	1988.134	10	3	29	57.669388	0.041	27	56	15.49998	0.53	−0.671
0332−403	P 0332−403	1983.633	49	3	34	13.654550	0.111	−40	8	25.39683	0.81	−0.882
0333+321	NRAO 140	1983.205	123	3	36	30.107639	0.029	32	18	29.34300	0.30	−0.838
0336−019	CTA 26	1985.726	89	3	39	30.937748	0.011	−1	46	35.80241	0.26	−0.361

Table 5 (contd)

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
0342+147	0342+147	1987.811	28	3	45	6.416506	0.034	14	53	49.55854	0.46	−0.821
0355+508	NRAO 150	1982.000	1847	3	59	29.747309	0.007	50	57	50.16193	0.07	−0.238
0400+258	CTD 26	1987.340	3	4	3	5.586178	0.180	26	0	1.50284	1.83	−0.973
0402−362	P 0402−362	1985.129	104	4	3	53.749821	0.038	−36	5	1.91249	0.41	−0.915
0406−127	0406−127	1989.466	4	4	9	5.769566	0.151	−12	38	48.14061	2.13	−0.983
0406+121	GC 0406+12	1984.235	96	4	9	22.008692	0.027	12	17	39.84791	0.42	−0.766
0409+229	P 0409+22	1989.663	13	4	12	43.666847	0.016	23	5	5.45394	0.36	−0.545
0420−014	P 0420−01	1988.604	6728	4	23	15.800686	0.002	−1	20	33.06435	0.05	−0.254
0420+417	VRO 41.04.01	1980.724	11	4	23	56.009869	0.202	41	50	2.71570	2.89	−0.458
0423+233	GC 0423+23	1989.671	11	4	26	55.734795	0.011	23	27	39.63453	0.31	−0.565
0425+048	P 0425+048	1987.655	26	4	27	47.570328	0.104	4	57	8.32905	1.47	−0.977
0426+273	0426+273	1989.666	14	4	29	52.960754	0.013	27	24	37.87780	0.42	−0.472
0430+052	3C 120	1981.230	141	4	33	11.095490	0.023	5	21	15.62231	0.77	−0.558
0434−188	P 0434−188	1985.211	100	4	37	1.482675	0.017	−18	44	48.61311	0.29	−0.715
0438−436	P 0438−43	1982.778	40	4	40	17.179842	0.116	−43	33	8.60147	0.78	−0.864
0440+345	0440+345	1989.178	13	4	43	31.635271	0.035	34	41	6.66321	0.41	−0.772
0446+112	P 0446+11	1989.025	20	4	49	7.671068	0.027	11	21	28.59826	0.48	−0.745
0451−282	P 0451−28	1983.885	18	4	53	14.646607	0.096	−28	7	37.32491	1.17	−0.962
0454−234	0454−234	1990.200	613	4	57	3.179160	0.006	−23	24	52.01993	0.26	−0.326
0458+138	P 0458+138	1989.351	10	5	1	45.270744	0.106	13	56	7.22257	1.38	−0.968
0459+060	GC 0459+06	1989.592	8	5	2	15.445878	0.139	6	9	7.49573	2.01	−0.984
0500+019	0500+019	1988.822	12	5	3	21.197086	0.034	2	3	4.67747	0.51	−0.911
0502+049	P 0502+049	1989.433	11	5	5	23.184755	0.094	4	59	42.72493	1.29	−0.987
0454+844	0454+844	1986.115	67	5	8	42.363876	0.121	84	32	4.54414	0.21	0.019
0506+101	P 0506+101	1988.770	20	5	9	27.457073	0.027	10	11	44.60062	0.42	−0.888
0507+179	P 0507+17	1987.671	25	5	10	2.369150	0.028	18	0	41.58187	0.43	−0.704
0511−220	P 0511−220	1989.307	6	5	13	49.114277	0.032	−21	59	16.09140	0.59	−0.794
0528+134	P 0528+134	1987.290	12691	5	30	56.416737	0.001	13	31	55.15025	0.04	−0.400
0537−441	P 0537−441	1985.159	134	5	38	50.361164	0.074	−44	5	8.93651	0.59	−0.912
0537−158	P 0537−158	1988.071	3	5	39	32.010201	0.067	−15	50	30.32208	3.36	−0.395
0536+145	0536+145	1988.410	28	5	39	42.366028	0.025	14	33	45.56190	0.36	−0.911
0544+273	0544+273	1987.942	17	5	47	34.148983	0.046	27	21	56.84298	0.73	−0.700
0552+398	DA 193	1986.888	25513	5	55	30.805620	0.002	39	48	49.16539	0.03	−0.395
0556+238	0556+238	1988.579	31	5	59	32.033134	0.014	23	53	53.92704	0.27	−0.553
0600+177	0600+177	1988.235	28	6	3	9.130286	0.029	17	42	16.81109	0.40	−0.913
0605−085	P 0605−08	1981.668	12	6	7	59.699158	0.105	−8	34	49.97785	2.02	−0.766

Table 5 (contd)

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
0611+131	0611+131	1989.633	5	6	13	57.692436	0.210	13	6	45.40760	2.82	−0.989
0642+214	3C 166	1988.667	15	6	45	24.099539	0.031	21	21	51.20209	0.48	−0.783
0657+172	0657+172	1987.192	23	7	0	1.525573	0.043	17	9	21.70139	0.80	−0.805
0722+145	P 0722+145	1989.203	19	7	25	16.807827	0.027	14	25	13.74644	0.39	−0.916
0723−008	DW 0723−00	1983.649	82	7	25	50.639908	0.041	−0	54	56.54408	0.71	−0.889
0727−115	P 0727−11	1988.978	2111	7	30	19.112436	0.003	−11	41	12.60000	0.07	−0.306
0735+178	P 0735+17	1985.148	82	7	38	7.393737	0.009	17	42	18.99918	0.21	−0.686
0736+017	P 0736+01	1989.726	16	7	39	18.033884	0.015	1	37	4.61811	0.26	−0.892
0738+313	OI 363	1984.907	17	7	41	10.703663	0.062	31	12	0.22568	0.54	−0.977
0742+103	DW 0742+10	1987.855	1044	7	45	33.059521	0.004	10	11	12.69277	0.12	−0.463
0743+259	GC 0743+25	1988.918	15	7	46	25.874182	0.032	25	49	2.13490	0.52	−0.708
0745+241	B2 0745+24	1985.049	50	7	48	36.109272	0.031	24	0	24.11141	0.65	−0.553
0748+126	P 0748+126	1983.493	68	7	50	52.045738	0.030	12	31	4.82853	0.57	−0.740
0754+100	P 0754+100	1989.022	14	7	57	6.642954	0.028	9	56	34.85172	0.50	−0.803
0805−077	P 0805−07	1987.825	8	8	8	15.535896	0.148	−7	51	9.88473	2.29	−0.998
0814+425	OJ 425	1987.175	257	8	18	15.999655	0.014	42	22	45.41528	0.18	−0.365
0823+033	P 0823+033	1986.860	294	8	25	50.338337	0.006	3	9	24.52047	0.13	−0.636
0827+243	B2 0827+24	1983.847	17	8	30	52.086189	0.035	24	10	59.82098	0.46	−0.943
0836+710	4C 71.07	1981.082	17	8	41	24.365801	0.252	70	53	42.17249	1.67	0.161
0851+202	OJ 287	1986.910	20630	(8	54	48.874920)	(0.000)	(20	6	30.64131)	(0.00)	(0.000)
0859−140	P 0859−14	1984.361	19	9	2	16.831095	0.129	−14	15	30.87777	1.60	−0.971
0859+470	OJ 499	1984.806	34	9	3	3.990097	0.068	46	51	4.13771	1.85	−0.237
0912+029	P 0912+029	1989.279	14	9	14	37.913448	0.044	2	45	59.24608	0.68	−0.975
0919−260	0919−260	1989.603	16	9	21	29.354017	0.149	−26	18	43.38887	3.74	−0.920
0920−397	P 0920−39	1987.932	6	9	22	46.417778	0.206	−39	59	35.06447	2.14	−0.976
0923+392	4C 39.25	1986.915	23656	9	27	3.013901	0.002	39	2	20.85215	0.03	0.010
0925−203	P 0925−203	1987.866	11	9	27	51.823792	0.160	−20	34	51.22625	2.19	−0.993
0953+254	OK 290	1980.571	86	9	56	49.875427	0.018	25	15	16.04907	0.60	−0.725
1012+232	1012+232	1989.027	20	10	14	47.065491	0.029	23	1	16.57087	0.39	−0.959
1022+194	GC 1022+19	1989.603	8	10	24	44.809582	0.026	19	12	20.41663	0.53	−0.774
1034−293	P 1034−293	1989.093	535	10	37	16.079710	0.009	−29	34	2.81403	0.14	−0.499
1038+064	OL 064.5	1984.281	64	10	41	17.162470	0.027	6	10	16.92338	0.72	−0.603
1042+071	P 1042+071	1989.195	12	10	44	55.911251	0.036	6	55	38.26318	0.66	−0.875
1044+719	1044+719	1988.317	11	10	48	27.619911	0.117	71	43	35.93801	0.66	−0.479
1055+018	P 1055+01	1988.883	2635	10	58	29.605196	0.002	1	33	58.82394	0.08	−0.424
1104−445	P 1104−445	1987.230	132	11	7	8.694243	0.083	−44	49	7.62063	0.49	−0.877

Table 5 (contd)

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
1111+149	GC 1111+14	1984.858	23	11	13	58.695099	0.041	14	42	26.95282	0.62	−0.974
1116+128	P 1116+12	1984.281	7	11	18	57.301376	0.043	12	34	41.71983	0.71	−0.955
1123+264	P 1123+26	1985.652	131	11	25	53.711910	0.010	26	10	19.97890	0.17	−0.741
1127−145	P 1127−14	1982.901	56	11	30	7.052641	0.064	−14	49	27.38946	0.99	−0.876
1128+385	GC 1128+38	1984.798	82	11	30	53.282635	0.023	38	15	18.54746	0.50	−0.242
1144+402	1144+402	1986.348	1082	11	46	58.297935	0.006	39	58	34.30446	0.07	−0.066
1144−379	P 1144−379	1986.151	211	11	47	1.370696	0.041	−38	12	11.02475	0.36	−0.902
1145−071	1145−071	1989.367	7	11	47	51.553970	0.025	−7	24	41.14088	0.71	−0.580
1148−001	P 1148−00	1983.107	23	11	50	43.870846	0.075	−0	23	54.20517	1.25	−0.891
1150+812	1150+812	1989.044	43	11	53	12.499044	0.142	80	58	29.15409	0.33	0.247
1156+295	GC 1156+29	1989.093	1331	11	59	31.833910	0.004	29	14	43.82693	0.06	−0.144
1219+285	ON 231	1980.571	113	12	21	31.690565	0.016	28	13	58.50143	0.61	−0.639
1222+037	P 1222+037	1983.849	60	12	24	52.421879	0.031	3	30	50.29426	0.60	−0.804
1226+023	3C 273	1985.978	11902	12	29	6.699723	0.002	2	3	8.59916	0.06	0.129
1243−072	1243−072	1989.997	9	12	46	4.232075	0.023	−7	30	46.57384	0.39	−0.834
1244−255	P 1244−255	1985.767	118	12	46	46.802166	0.037	−25	47	49.29095	0.44	−0.922
1253−055	3C 279	1987.211	710	12	56	11.166497	0.005	−5	47	21.52418	0.16	0.153
1302−102	P 1302−102	1989.652	15	13	5	33.014919	0.019	−10	33	19.42628	0.36	−0.738
1308+326	B2 1308+32	1987.403	2312	13	10	28.663879	0.004	32	20	43.78286	0.06	−0.205
1313−333	OP−322	1984.363	45	13	16	7.985952	0.093	−33	38	59.17341	1.00	−0.947
1315+346	OP 326	1990.151	4	13	17	36.494152	0.035	34	25	15.93075	1.46	0.839
1335−127	DW 1335−12	1989.285	1379	13	37	39.782745	0.004	−12	57	24.69304	0.10	−0.148
1342+662	GC 1342+662	1983.849	35	13	43	45.959625	0.062	66	2	25.74482	0.44	−0.053
1342+663	GC 1342+663	1984.708	162	13	44	8.679707	0.034	66	6	11.64348	0.22	0.150
1349−439	P 1349−439	1985.833	45	13	52	56.535149	0.143	−44	12	40.39031	0.90	−0.890
1354+195	P 1354+19	1986.707	739	13	57	4.436650	0.006	19	19	7.37252	0.11	−0.396
1354−152	OP−192	1989.104	18	13	57	11.245065	0.029	−15	27	28.78804	0.38	−0.913
1404+286	OQ 208	1986.855	16207	14	7	0.394422	0.002	28	27	14.68967	0.05	−0.291
1406−076	P 1406−076	1989.701	7	14	8	56.481178	0.034	−7	52	26.66544	0.54	−0.765
1418+546	GC 1418+54	1985.652	531	14	19	46.597424	0.014	54	23	14.78695	0.15	−0.163
1424−418	P 1424−41	1988.396	5	14	27	56.297581	0.132	−42	6	19.44042	0.87	−0.833
1430−178	OQ−151	1983.981	23	14	32	57.690651	0.273	−18	1	35.25099	3.75	−0.974
1443−162	1443−162	1989.805	6	14	45	53.376358	0.093	−16	29	1.61882	1.25	−0.956
1445−161	P 1445−16	1989.518	12	14	48	15.054192	0.035	−16	20	24.54957	0.50	−0.908
1502+106	OR 103	1988.423	2420	15	4	24.979795	0.003	10	29	39.19889	0.10	−0.081
1504−166	P 1504−167	1984.760	64	15	7	4.786958	0.022	−16	52	30.26719	0.63	−0.523

Table 5 (contd)

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
1510−089	P 1510−08	1986.814	244	15	12	50.532932	0.013	−9	5	59.82971	0.26	−0.707
1511−100	P 1511−100	1989.405	9	15	13	44.893406	0.027	−10	12	0.26344	0.58	−0.671
1514−241	P 1514−24	1989.545	10	15	17	41.813305	0.058	−24	22	19.47806	0.70	−0.952
1519−273	P 1519−273	1986.975	180	15	22	37.676056	0.049	−27	30	10.78650	0.57	−0.938
1532+016	P 1532+01	1988.806	16	15	34	52.453609	0.034	1	31	4.20772	0.53	−0.908
1548+056	DW 1548+05	1986.819	2423	15	50	35.269226	0.003	5	27	10.44892	0.09	0.113
1555+001	DW 1555+00	1985.066	96	15	57	51.433945	0.025	−0	1	50.41273	0.61	−0.651
1611+343	DA 406	1989.841	1744	16	13	41.064259	0.003	34	12	47.90900	0.08	−0.171
1614+051	P 1614+051	1988.544	13	16	16	37.556808	0.028	4	59	32.73700	0.45	−0.875
1622−253	P 1622−253	1990.195	525	16	25	46.891602	0.007	−25	27	38.32558	0.28	−0.341
1633+382	GC 1633+38	1988.861	5684	16	35	15.492970	0.003	38	8	4.50046	0.06	−0.070
1637+574	P 1637+574	1983.521	271	16	38	13.456281	0.021	57	20	23.97837	0.24	0.038
1638+398	NRAO 512	1983.405	119	16	40	29.632765	0.021	39	46	46.02849	0.30	−0.599
1642+690	1642+690	1980.699	1079	16	42	7.848453	0.015	68	56	39.75576	0.10	−0.060
1641+399	3C 345	1985.825	19307	16	42	58.809959	0.002	39	48	36.99369	0.06	−0.008
1655+077	OS 092	1987.989	27	16	58	9.011436	0.025	7	41	27.54106	0.71	−0.579
1656+053	DW 1656+05	1982.948	20	16	58	33.447254	0.117	5	15	16.44569	1.76	−0.992
1657−261	P 1657−261	1985.153	39	17	0	53.154152	0.046	−26	10	51.72603	0.63	−0.855
1706−174	OT−111	1984.986	41	17	9	34.345371	0.048	−17	28	53.36405	0.77	−0.881
1717+178	GC 1717	1981.466	25	17	19	13.048463	0.030	17	45	6.43817	1.60	0.058
1730−130	NRAO 530	1987.879	1156	17	33	2.705772	0.005	−13	4	49.54746	0.14	−0.183
1738+476	OT 465	1985.490	93	17	39	57.129055	0.023	47	37	58.36189	0.57	−0.243
1739+522	4C 51.37	1989.997	2728	17	40	36.977840	0.004	52	11	43.40747	0.06	0.066
1741−038	P 1741−038	1988.664	5917	17	43	58.856140	0.002	−3	50	4.61590	0.10	0.135
1748+701	1749+701	1982.658	138	17	48	32.840473	0.064	70	5	50.76680	0.27	−0.053
1749+096	OT 081	1987.660	1447	17	51	32.818584	0.003	9	39	0.72882	0.10	0.061
1803+784	1803+784	1987.488	23352	18	0	45.683819	0.009	78	28	4.01828	0.04	0.075
1807+698	3C 371	1984.948	238	18	6	50.680634	0.038	69	49	28.10885	0.20	−0.021
1821+107	P 1821+10	1985.019	45	18	24	2.855270	0.034	10	44	23.77293	1.38	−0.523
1845+797	3C 390.3	1979.901	16	18	42	8.989495	0.362	79	46	17.12646	0.58	−0.118
1908−201	OV−213	1986.907	32	19	11	9.652929	0.035	−20	6	55.10953	0.47	−0.901
1920−211	OV−235	1988.276	37	19	23	32.189861	0.030	−21	4	33.33339	0.43	−0.897
1921−293	OV−236	1988.921	1040	19	24	51.055946	0.006	−29	14	30.12061	0.15	−0.252
1923+210	OV 239.7	1987.244	88	19	25	59.605373	0.016	21	6	26.16149	0.26	−0.799
1928+738	1928+738	1981.460	51	19	27	48.494961	0.089	73	58	1.57013	0.71	−0.293
1936−155	P 1936−15	1988.959	16	19	39	26.657772	0.031	−15	25	43.05868	0.48	−0.886

Table 5 (contd)

IAU name	Common name	Mean observation epoch	Number of observations	Right ascension			Error, msec	Declination			Error, mas	RA, dec. correlation
				hr	min	sec		d	m	asec		
1954+513	1954+513	1989.238	24	19	55	42.738186	0.043	51	31	48.54832	0.48	0.026
1958−179	OV−198	1986.285	119	20	0	57.090458	0.023	−17	48	57.67260	0.33	−0.855
2008−159	P 2008−159	1988.844	20	20	11	15.710927	0.027	−15	46	40.25305	0.42	−0.868
2021+614	OW 637	1987.616	101	20	22	6.681649	0.032	61	36	58.80481	0.50	−0.009
2030+547	OW 551	1981.630	14	20	31	47.958549	0.197	54	55	3.14788	3.01	−0.080
2029+121	P 2029+121	1984.093	33	20	31	54.994236	0.062	12	19	41.33965	1.26	−0.715
2037+511	3C 418	1988.150	359	20	38	37.034779	0.010	51	19	12.66291	0.11	0.217
2113+293	B2 2113+29B	1986.307	178	21	15	29.413473	0.010	29	33	38.36657	0.23	−0.162
2121+053	OX 036	1989.107	5908	21	23	44.517380	0.002	5	35	22.09376	0.08	0.113
2128−123	P 2128−12	1988.831	21	21	31	35.261684	0.025	−12	7	4.79540	0.41	−0.849
2131−021	P 2131−021	1988.885	25	21	34	10.309592	0.016	−1	53	17.23844	0.32	−0.759
2134+004	P 2134+004	1985.852	6648	21	36	38.586308	0.002	0	41	54.21443	0.10	−0.111
2145+067	P 2145+06	1986.433	2454	21	48	5.458668	0.003	6	57	38.60450	0.08	0.102
2149+056	OX 082	1984.142	78	21	51	37.875381	0.039	5	52	12.95605	0.63	−0.873
2155−152	OX−192	1985.260	51	21	58	6.281964	0.089	−15	1	9.32834	1.26	−0.961
2200+420	VRO 42.22.01	1986.403	15662	22	2	43.291370	0.003	42	16	39.98008	0.04	0.405
2201+315	B2 2201+31A	1983.433	24	22	3	14.975830	0.038	31	45	38.27049	0.70	0.170
2216−038	P 2216−03	1987.052	2789	22	18	52.037716	0.003	−3	35	36.87867	0.10	0.068
2223−052	3C 446	1985.274	68	22	25	47.259274	0.018	−4	57	1.39028	0.40	−0.687
2227−088	P 2227−08	1988.855	19	22	29	40.084297	0.026	−8	32	54.43421	0.43	−0.891
2229+695	2229+695	1988.423	16	22	30	36.469852	0.092	69	46	28.07718	0.39	0.065
2230+114	CTA 102	1985.129	125	22	32	36.408909	0.013	11	43	50.90439	0.24	−0.734
2234+282	GC 2234+28	1988.735	3182	22	36	22.470861	0.003	28	28	57.41359	0.06	0.197
2243−123	OY−172.6	1985.112	121	22	46	18.231953	0.019	−12	6	51.27657	0.32	−0.826
2245−328	P 2245−328	1984.167	65	22	48	38.685780	0.045	−32	35	52.18752	0.49	−0.914
2251+158	3C 454.3	1986.145	13876	22	53	57.747932	0.002	16	8	53.56100	0.05	0.315
2253+417	GC 2253+41	1983.351	54	22	55	36.707783	0.043	42	2	52.53282	0.97	0.032
2254+074	GC 2254+07	1989.614	4	22	57	17.303083	0.159	7	43	12.30326	2.56	−0.962
2254+024	P 2254+024	1988.749	19	22	57	17.563122	0.047	2	43	17.51163	0.70	−0.945
2255−282	P 2255−282	1990.205	405	22	58	5.962849	0.008	−27	58	21.25589	0.33	−0.183
2318+049	GC 2318+04	1988.970	24	23	20	44.856579	0.025	5	13	49.95361	0.38	−0.904
2320−035	P 2320−035	1985.216	89	23	23	31.953726	0.024	−3	17	5.02289	0.40	−0.885
2335−027	P 2335−027	1989.570	8	23	37	57.339039	0.069	−2	30	57.62940	1.09	−0.917
2345−167	P 2345−16	1984.839	71	23	48	2.608478	0.038	−16	31	12.02106	0.56	−0.898
2351−154	2351−154	1989.367	9	23	54	30.195244	0.071	−15	13	11.21327	1.02	−0.948
2355−106	P 2355−106	1987.701	40	23	58	10.882395	0.033	−10	20	8.61091	0.55	−0.844

Table 6. JPL 1990-3 celestial ephemeris pole motion model

Index	Period, days	ΔA_ψ (in phase) (out of phase), nrad	ΔA_ϵ (in phase) (out of phase), nrad
1	6798.38	−22.1±1.7 7.7 0.7	8.2±0.3 10.3 0.2
2	3399.19	7.8±0.6 1.0 0.3	0.6±0.1 0.7 0.1
10	365.26	24.6±0.1 5.5 0.1	9.3±0.1 −0.9 0.1
9	182.62	6.9±0.1 −6.1 0.1	−2.4±0.1 −1.9 0.1
11	121.75	−1.0±0.1 0.1 0.1	0.8±0.1 0.1 0.1
32	27.55	−1.0±0.1 −0.6 0.1	−0.5±0.1 0.4 0.1
31	13.66	−2.7±0.2 0.6 0.2	1.4±0.1 −0.4 0.1
33	13.63	−2.0±0.2 0.1 0.2	0.6±0.1 0.3 0.1
—	433.20	−3.0±0.1 3.0 0.1	−0.6±0.1 −1.3 0.1
Precession, nrad/yr		−10.4±0.3	—
Offset		−61.3±1.6	9.4±0.3

Table 7. Comparisons of other JPL catalogs with JPL 1990-3

Catalog	1987-1	1991-1
Number of common sources	103	208
RMS uncertainty for common sources		
JPL 1990-3: RAcos δ and dec., nrad	5.0, 4.4	6.3, 5.4
Other: RAcos δ and dec., nrad	9.0, 8.6	4.8, 4.0
Rotational offsets, nrad x	−2.8	0.0
y	7.2	1.8
z	−3.1	0.5
χ^2 per degree of freedom	1.7	1.5
RMS difference, nrad: RAcos δ	7.5	5.2
dec.	10.8	5.8
χ^2 per degree of freedom: RAcos δ	2.0	1.8
dec.	1.3	1.1

Table 8. Comparisons of external catalogs with JPL 1990-3

Catalog	IERS90	GSFC89	GSFC90
Number of common sources	183	120	159
RMS uncertainty for common sources			
JPL 1990-3: RAcos δ and dec., nrad	4.9, 4.3	4.7, 3.6	6.1 4.8
Other: RAcos δ and dec., nrad	7.6, 6.8	5.7, 3.0	3.7 2.9
Rotational offsets, nrad x	1.4	10.2	−5.4
y	0.7	5.0	−11.9
z	0.3	−24.3	1.8
χ^2 per degree of freedom	3.6	2.7	3.8
RMS difference, nrad: RAcos δ	5.8	4.2	7.5
dec.	8.9	6.2	6.6
χ^2 per degree of freedom: RAcos δ	3.3	2.1	4.0
dec.	4.0	2.7	3.3

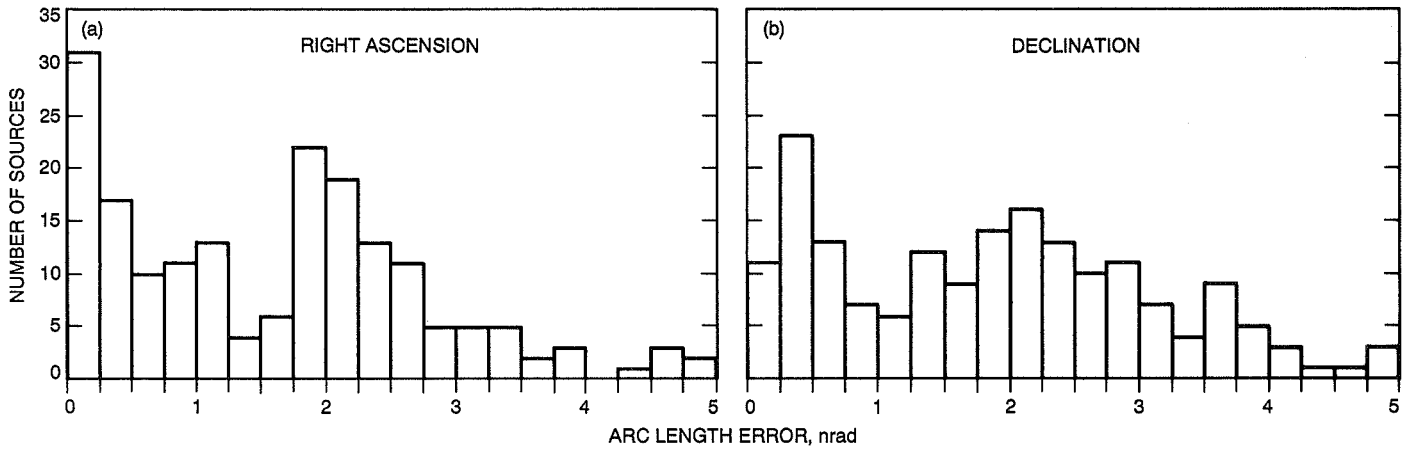


Fig. 1. Distribution of errors in (a) right ascension (arc length) and (b) declination for source coordinates in the catalog JPL 1990-3.

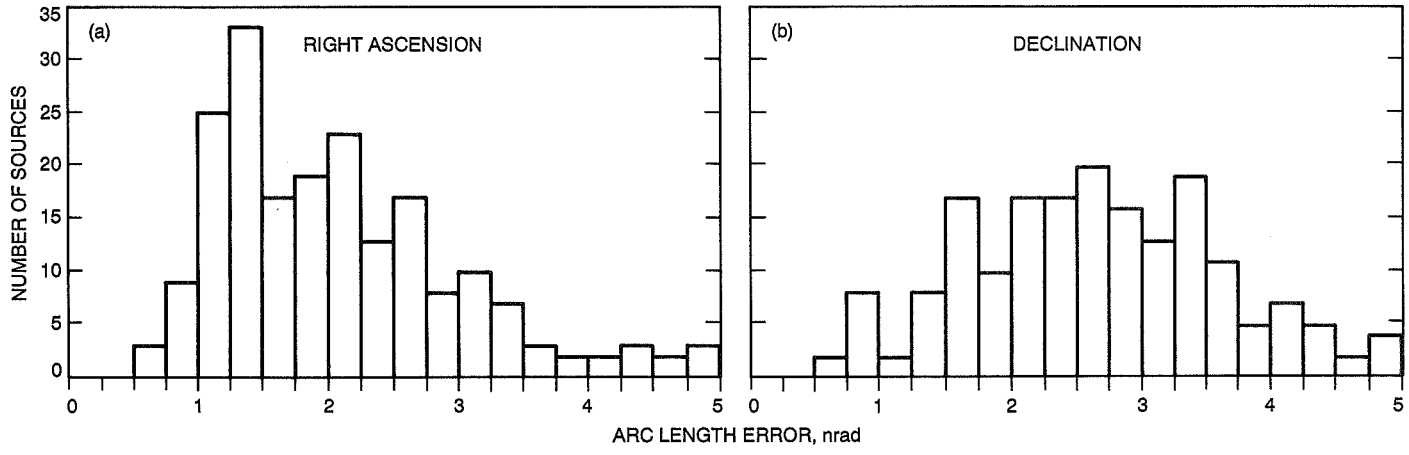


Fig. 2. Distribution of errors in (a) right ascension (arc length) and (b) declination for source coordinates in the TEMPO catalog JPL 1991-1.

Wideband Waveguide Polarizer Development for SETI

P. Lee and P. Stanton
Ground Antennas and Facilities Engineering Section

A wideband polarizer for the Deep Space Network (DSN) 34-meter beam-waveguide antenna is needed for the Search for Extraterrestrial Intelligence (SETI) project. The results of a computer analysis of a wideband polarizer are presented in this article.

I. Introduction

The Search for Extraterrestrial Intelligence (SETI) sky survey will scan a frequency range from 1 to 10 GHz, using the Deep Space Network (DSN) 34-meter beam-waveguide antennas. To minimize the number of feeds that accommodate this broad frequency range, wideband components are necessary. In [1], a design for a wideband quarter-wave polarizer was presented. The polarizer consists of a rectangular waveguide with all four walls loaded with natural or artificial dielectrics. This article describes the performance of such a polarizer, as analyzed with Hewlett-Packard's 85180A High-Frequency Structure Simulator (HFSS), a software package that uses the finite-element method to analyze closed structures.

II. Method

Figure 1 shows the geometry of the polarizer. An incident TE_{10} mode excites a quasi-TE_{10} mode in the loaded waveguide. An incident TE_{01} mode excites a quasi-TE_{01} mode, which has a different propagation constant from the quasi-TE_{10} mode. As the two modes travel through the loaded waveguide, the phase shift between them is proportional to the difference between the propagation constants. Since the propagation constants are functions of frequency, the phase shift is also. It was found in [1] that by loading the four walls of the polarizer with dielectric, the differential phase shift can be held constant to within a given tolerance over a wider bandwidth than in polarizers where only two opposing walls are loaded.

The design parameters were chosen by assuming that the forms of the quasi-TE_{10} and quasi-TE_{01} modes in the loaded waveguide did not vary much from the TE_{10} and TE_{01} modes in an unloaded waveguide. Since the fields of the TE_{10} and TE_{01} modes are zero on the sidewalls parallel to the E-field components, and since the dielectric on the walls is thin, only the dielectric on the walls perpendicular to the E-field was considered when each mode was analyzed for the initial design. This simpler problem of a waveguide with only two walls loaded has been solved [2].

The polarizer was analyzed on a Sun SPARC 1+ workstation using HFSS software, revision A.01.00. The workstation has approximately 500 MB of hard-disk memory

available in the user partition. The HFSS software analyzes arbitrary closed structures with finite-element methods. It computes the eigenmodes at each port by using the number of modes chosen by the user. Then the S-parameters for the structure are calculated to within a tolerance set by the user. The finite-element mesh is refined over a number of passes until the desired tolerance is achieved. Once a mesh is created at one frequency, the same mesh can be used in a sweep over nearby frequencies. After HFSS calculates the S-parameters, it calculates the impedances at each port.

The first step in analyzing the polarizer was to determine the propagation constants of the quasi-TE_{10} and quasi-TE_{01} modes in the loaded waveguide. A 0.25-in. long slice of the waveguide was used. Electric and magnetic planes of symmetry were employed to reduce the cross-section to one quarter of its original size. By placing the planes of symmetry as shown in Fig. 2(a), the mode resulting from an incident TE_{10} mode could be analyzed. Similarly, Fig. 2(b) shows the placement of the planes of symmetry used to analyze the mode resulting from an incident TE_{01} mode. An adaptive pass was run at 11.0 GHz to create the finite-element mesh, and then a frequency sweep was performed from 6.0 GHz to 16.0 GHz in 0.5-GHz intervals. Adaptive passes were later run at the endpoint frequencies to check the validity of the mesh. The frequency range was chosen for measurement convenience and will be scaled to the SETI frequency range later.

The polarizer could not be analyzed as one structure because this presented too large a problem for the memory in the workstation. An interface section consisting of a thin slice of the loaded waveguide and a thin slice of empty waveguide, as shown in Fig. 3(a), was analyzed instead. The HFSS software assumes that each port is connected to an infinitely long structure having the same cross-section as the port, so the effects of the interface by itself can be deduced from this structure. The S-parameters for this structure were calculated over the frequency range 9.5 GHz to 15.5 GHz. HFSS has the capability of computing the S-parameters resulting from the addition of an extension to any port if the extension has the same cross-sectional geometry as the port. By using this feature, the S-parameters for the structure shown in Fig. 3(b) were computed. Finally, the S-parameters from the structures shown in Figs. 3(a) and 3(b) were combined to create the S-parameters for the entire polarizer, as shown in Fig. 3(c).

III. Results

Figure 4(a) shows the computed propagation constants of the quasi-TE_{10} and quasi-TE_{01} modes in the loaded waveguide versus frequency. The difference between the propagation constants, shown in Fig. 4(b), varies less than 1 percent between 11 GHz and 15.5 GHz. Running one adaptive pass on the HFSS software at 11.0 GHz and then sweeping over the entire frequency range seemed to give valid answers. The propagation constants calculated from the sweep varied a maximum of 0.01 percent from the propagation constants calculated from adaptive passes run at the endpoints.

In an empty waveguide, the eigenmodes on the surface of the input ports are independent of frequency; however, with the dielectric loading, this is no longer true [2]. Figure 5 shows the quasi-TE_{10} and quasi-TE_{01} eigenmodes at the ports for (a) a loaded waveguide at 6.0 GHz, (b) a loaded waveguide at 16.0 GHz, and (c) an unloaded waveguide. As the frequency increases, the modes in the loaded waveguide vary more from the TE_{10} and TE_{01} modes in an empty waveguide. At 16.0 GHz, sizable fields exist in the dielectric strips on the walls parallel to the fields of the empty waveguide modes.

With the eigenmodes dependent on frequency, some complications occur which the HFSS software does not handle well when a frequency sweep is performed. The eigenmodes are computed at each frequency and then ordered according to the propagation constants. The quasi-TE_{10} and quasi-TE_{01} eigenmodes in the loaded waveguide are ordered as the first and second modes at 9.5 GHz, but then are ordered as the first and third modes from somewhere around 10 GHz and above. When analyzing the interface section, three modes were used to ensure that both the modes of interest were always included. The HFSS only plots the eigenmodes for the last frequency computed, so the user does not necessarily know what the mode ordering is at lower frequencies in the sweep. Also, HFSS will not compute impedances unless the user has defined impedance lines on the ports. The impedance lines are the lines connecting the points of maximum voltage difference on the ports, and are different for each mode. Only one set of impedance lines may be defined for all the frequencies in a sweep, and they must be defined in the same order as the corresponding modes. When the modes are switched during a sweep, the impedance lines are no longer valid at every frequency, and impedances will not be computed for any modes at any frequencies.

The S-parameters for the interface section between a 0.01-in. long slice of unloaded waveguide and a 0.1-in. long slice of loaded waveguide are plotted in Fig. 6. These S-parameters were used to construct the S-parameters for an entire polarizer of length 4.73 in. The S-parameters and differential phase shift for this polarizer are shown in

Fig. 7. The magnitude of the S_{11} parameter is large, indicating a poor impedance match at the interface. Also, the differential phase shift is not very constant, which also may be caused by a poor impedance match at the interface. To correct these problems, an input-matching section needs to be designed.

In an initial attempt to design an input-matching section, the impedances given by HFSS were examined. HFSS computes the impedance three ways. The first impedance, Z_{pi}, is derived from the input power and current calculated in the field solution. The second impedance, Z_{pv}, is computed from the input power from the field solution and the voltage across the impedance line on the port (the impedance line is defined by the user). The third impedance, Z_{vi}, is the geometric mean of the other two. These three impedances should be the same except for a multiplying constant; however, this was not the case. The three impedances, normalized to the 9.5-GHz values, are compared in Fig. 8.

IV. Conclusions

The results shown in Fig. 4 indicate that the polarizer design will function over a wide band. However, the return loss of the polarizer is quite high, and the differential phase shift is not as constant as desired. An input-matching section needs to be designed to correct both of these problems. The impedances as calculated by HFSS seem to be inconsistent. The values for S-parameters and impedances given by HFSS need to be compared with actual measurements on a polarizer to verify the results.

The HFSS software on the Sun SPARC 1+ workstation runs too slowly and uses too much hard-disk memory to handle large problems. As the number of modes, the size of the structure, and the complexity of the structure increase, HFSS rapidly slows down and uses more memory. Whenever possible, the problem needs to be broken down into smaller, more manageable pieces which can be recombined after analysis with HFSS.

References

[1] E. Lier and T. Schaug-Pettersen, "A Novel Type of Waveguide Polarizer with Large Cross-Polar Bandwidth," *IEEE Transactions on Microwave Theory and Techniques*, vol. 36, pp. 1531–1534, November 1988.

[2] R. F. Harrington, *Time Harmonic Electromagnetic Fields*, New York: McGraw-Hill, pp. 158–163, 1961.

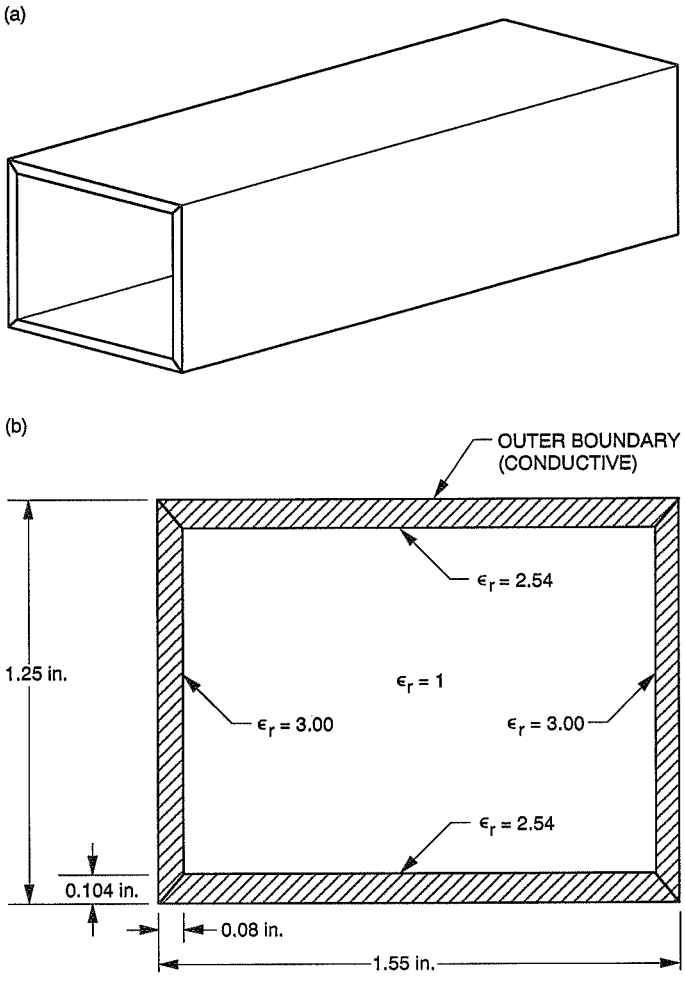


Fig. 1. Geometry of the polarizer: (a) polarizer and (b) cross-section.

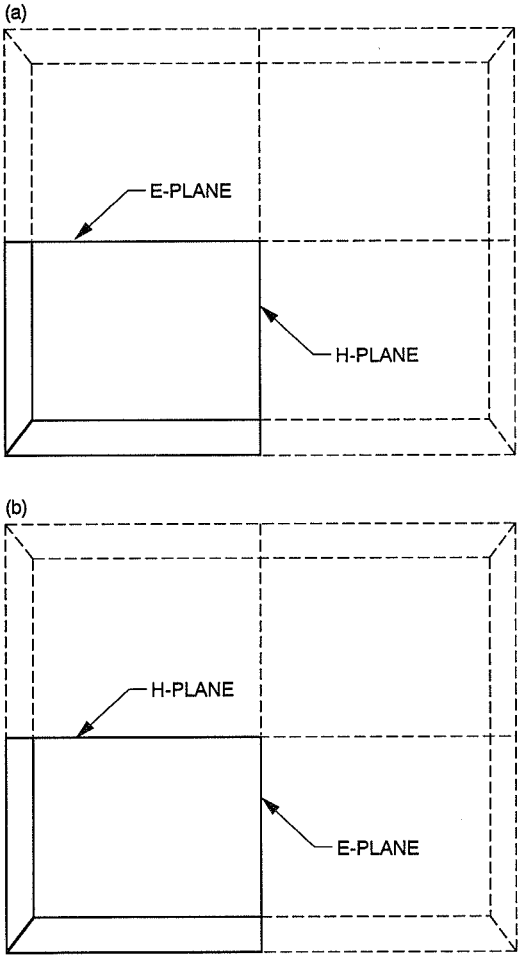


Fig. 2. Placement of planes of symmetry: (a) to analyze mode resulting from a TE_{10} mode input; and (b) to analyze mode resulting from a TE_{01} mode input.

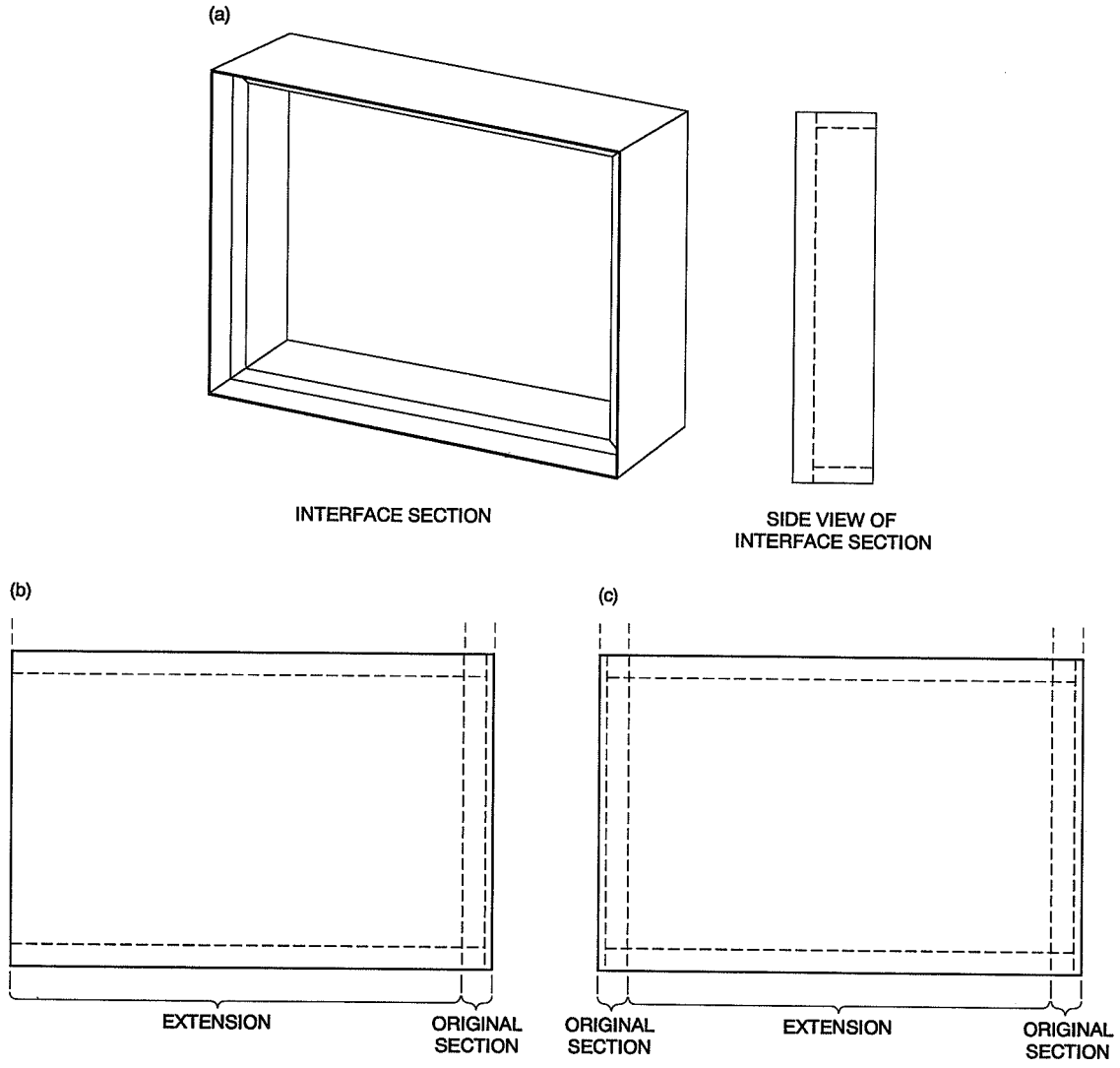


Fig. 3. Stages of polarizer analysis: (a) interface section consisting of a thin slice of the loaded waveguide and a thin slice of empty waveguide; (b) side view of original section with extension; and (c) side view of total polarizer.

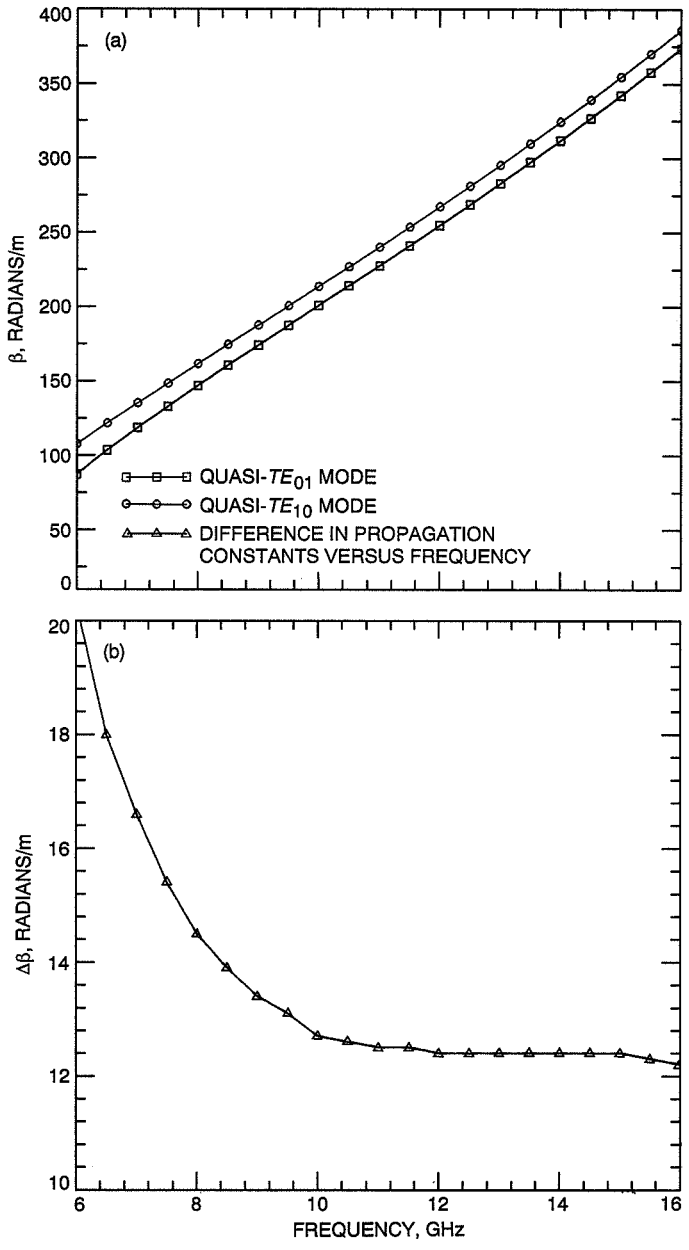


Fig. 4. Propagation constants of quasi-TE_{10} and quasi-TE_{01} modes: (a) β versus frequency and (b) $\Delta\beta$ versus frequency.

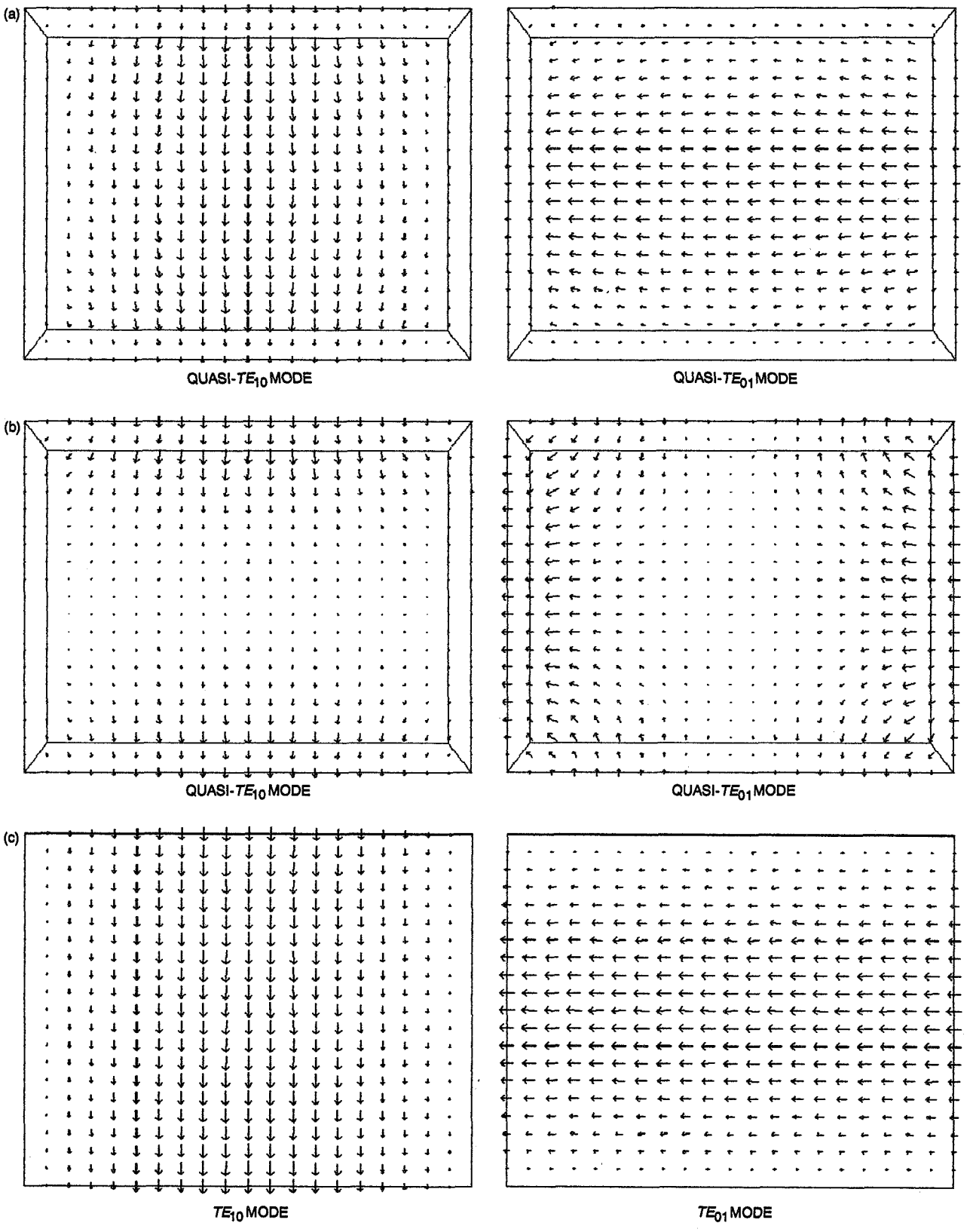


Fig. 5. Eigenmodes of interest at ports for (a) a loaded waveguide at 6.0 GHz; (b) a loaded waveguide at 16.0 GHz; and (c) an unloaded waveguide.

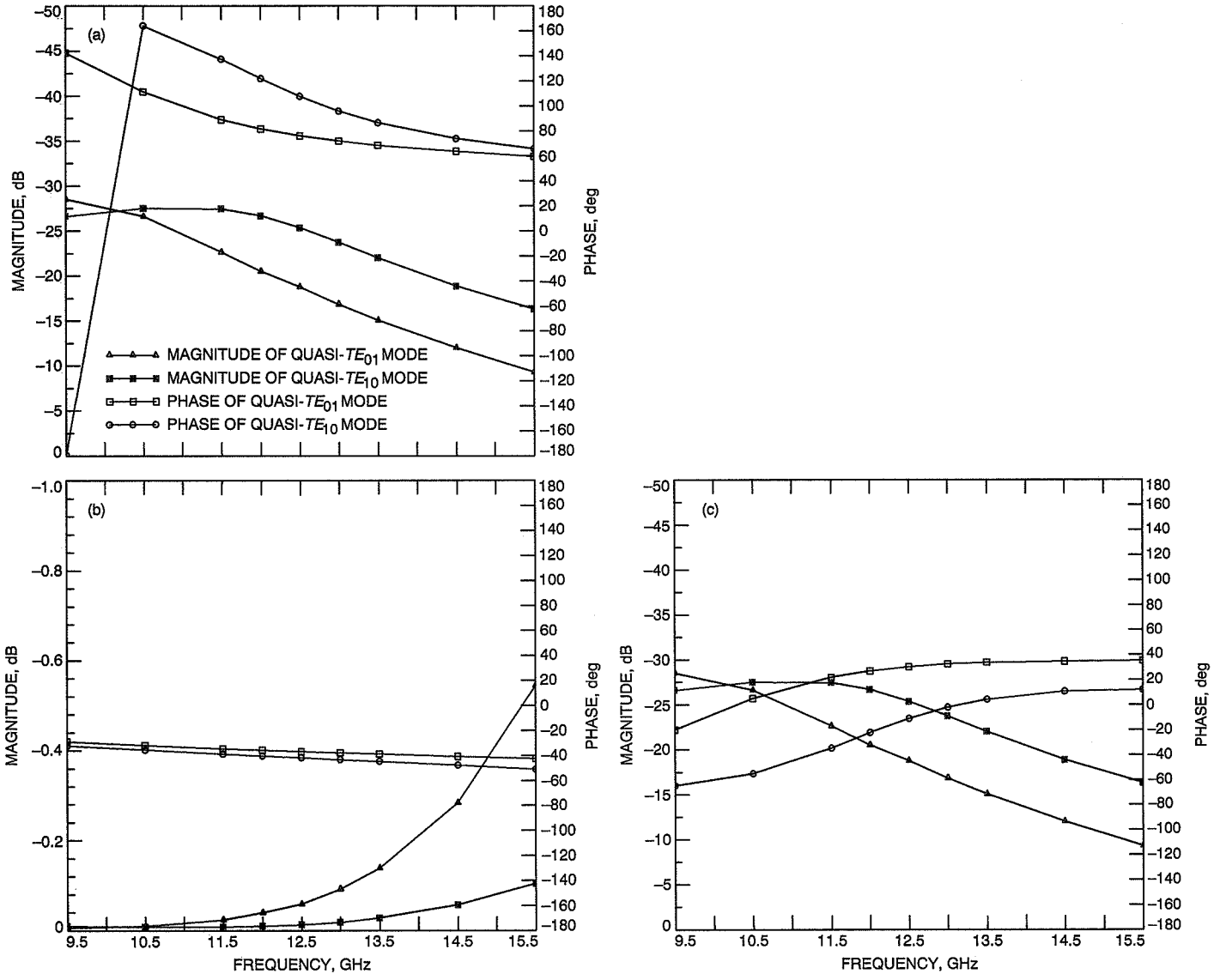


Fig. 6. *S*-parameters for interface section: (a) S_{11}; (b) S_{12} (= S_{21}); and (c) S_{22}.

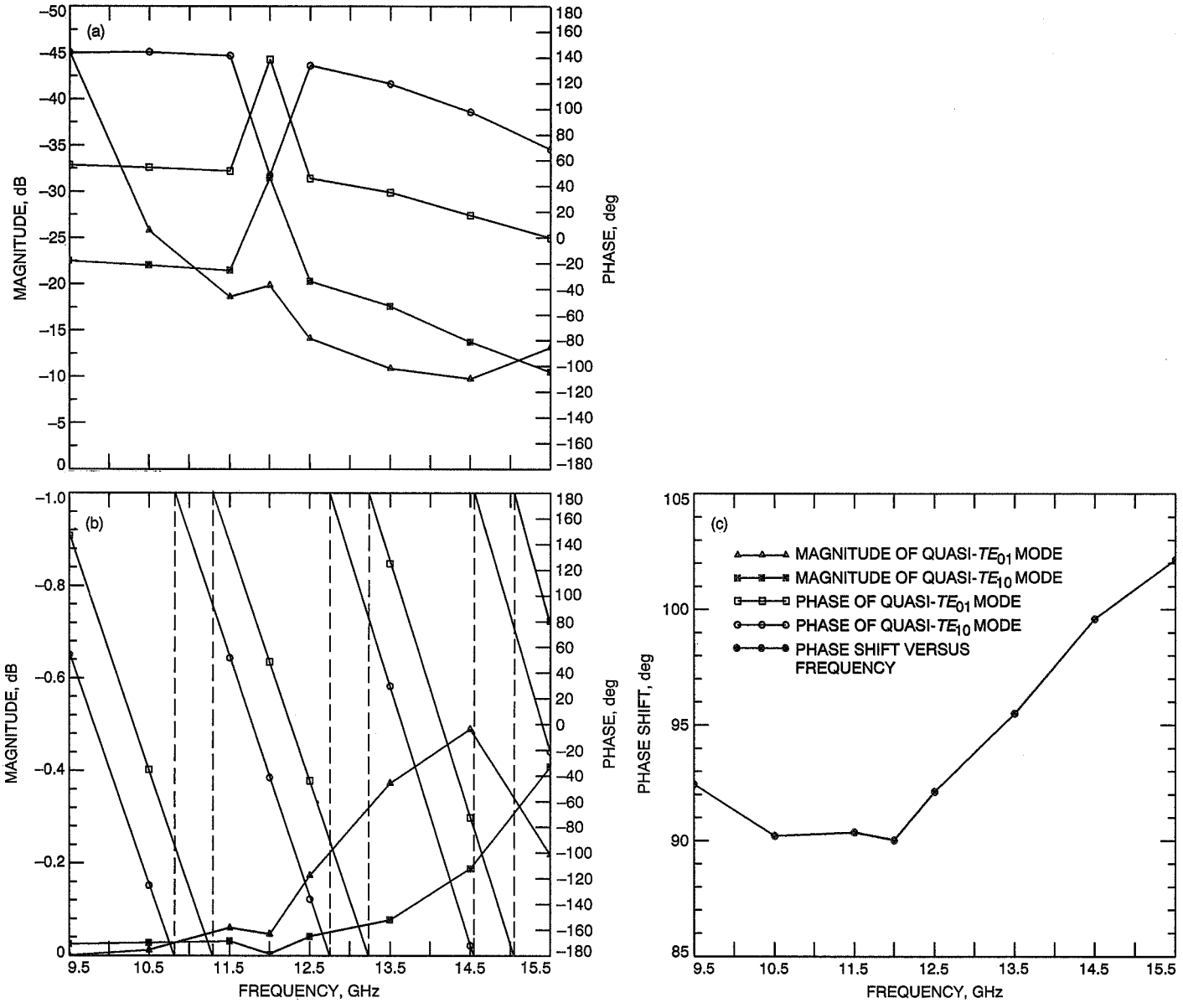


Fig. 7. Polarizer data: (a) S_{11} parameter; (b) S_{12} parameter; and (c) phase shift.

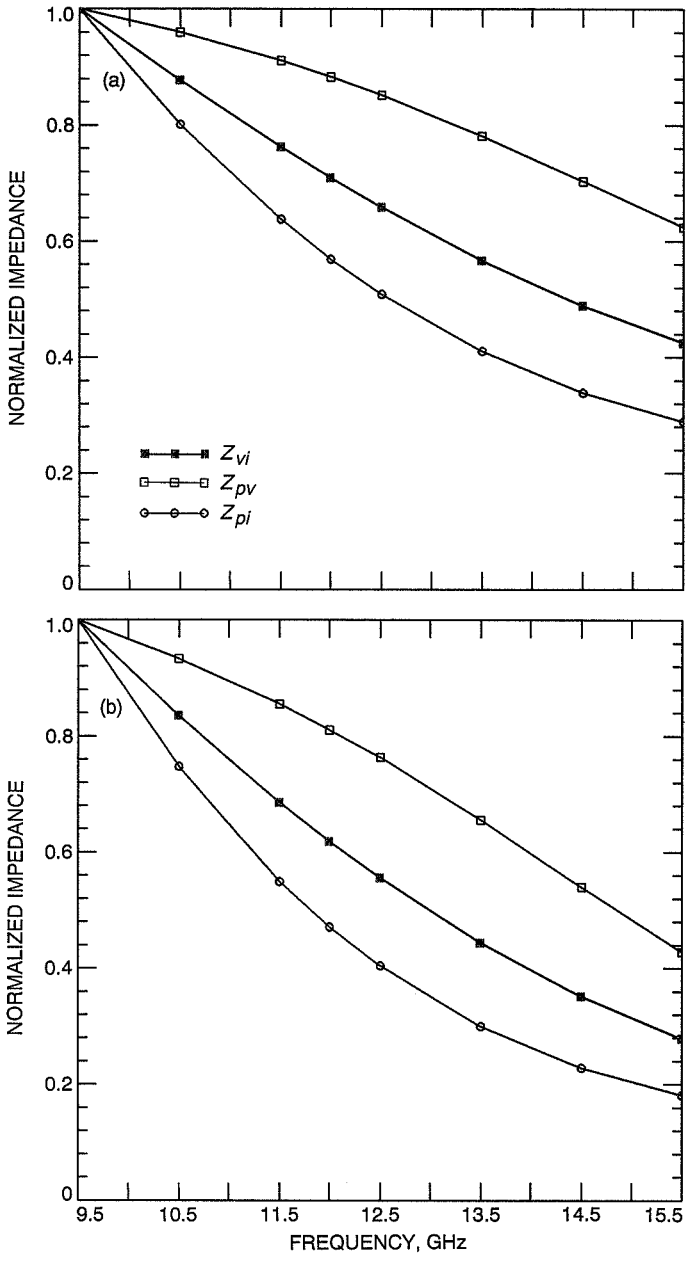


Fig. 8. Normalized impedances for (a) quasi-TE_{10} mode and (b) quasi-TE_{01} mode.

Stochastic Availability Analysis of Operational Data Systems in the Deep Space Network

T. N. Issa*
TDA Mission Support and DSN Operations

Existing availability models of standby redundant systems consider only an operator's performance and its interaction with the hardware performance. In the case of operational data systems in the DSN, in addition to an operator–system interface, a controller reconfigures the system and links a standby unit into the network data-path upon failure of the operating unit. In this article, a stochastic (Markovian) process technique is employed to model and analyze the availability performance of this class of communication systems. The link-controller's performance and occurrence of degradation due to partial failures are quantitatively incorporated into the model. Exact expressions of the steady-state availability and proportion-degraded performance measures are derived for the systems under study. The interaction among the hardware, operator, and controller performance parameters and that interaction's effect on data availability are evaluated and illustrated for an operational data-processing system.

I. Introduction

Extensive research and practical literature exists in relation to the analysis of human performance and reliability modeling of man–machine systems. An up-to-date list and classification of the body of literature addressing human errors and their effect on the reliability and availability of engineering systems are provided in [1]. Human error is defined as a failure to perform a prescribed act (or an out-of-tolerance human action) that could result in damage to equipment or the disruption of scheduled operations [2,3]. Categories and causal analysis of human errors are detailed in [4–6].

Human performance is not easy to quantify and predict. Rigby [7] stated that error is a function of human variability; however, there are several other factors that may be considered in the analysis and quantitative measurement of human performance [8–10]. The problem of human reliability prediction has been discussed earlier by Meister [11] and Regulinski [12]. Recent studies, such as those of Dhillon [13–17] and Gupta [18,19] have emphasized reliability and availability modeling of standby systems with human-initiated failures. In these efforts, the Markov pro-

*Bendix Field Engineering Corporation

cess was employed to model the interaction of hardware and human performances. In these models, hardware units of only two states (operating and failed) were considered, and the impact of one type of human interface was incorporated, that is, the performance of operating personnel only. One assumption common to these models is that the switch over to a standby unit (i.e., restoration of system operation) is instantaneous and perfect.

In the telecommunication and data-processing systems implemented in the ground-based communication complexes of the DSN, there are two types of human–equipment interface. One type of interface involves an operator's monitor and control of an operational unit, and the second type requires a link controller's interface with a console display to perform system reconfiguration and restoration by linking a standby unit. Operator errors may cause failure (disruption of operation) or performance degradation of the operating unit. Errors made by the link controller would usually increase the completion time of the reconfiguring and linking task.

In this study, the stochastic operational behavior of two-unit standby data-processing systems at the DSN is investigated. In addition to operator-initiated outages, these systems are characterized by a noninstantaneous restoration of their operation through a controller-performed linking process. The purpose of this analysis is to apply a Markov process technique to develop steady-state expressions for "operational" availability and proportion-degraded performance estimation and to implement them as relevant measures of performance for the systems under study. Of particular significance is the evaluation of the impact of various levels of link-controller performance on the proposed performance measures at selected levels of operator error rate.

II. The Human Role in Data-Processing Systems Operation and Maintenance

In this study, a class of communication systems utilized during spacecraft tracking and data acquisition is defined. This class includes such systems as the receiver, telemetry processing, command assemblies, and others. A typical two-unit standby configuration used in ground-based stations is considered representative of these systems for reliability and availability modeling.

In any configuration, an "operational" unit or assembly is viewed as a complex structure of hardware and human components. The human components and their activities include:

1. *Operator:* Performs time-continuous operation and monitoring of equipment. The operator provides inputs and responses, performs tests and verifications, and monitors the operating unit.

2. *Link controller:* Performs system reconfiguration and standby-unit linking upon failure of the operating unit. The link controller interfaces with a console display to monitor data flow and link status parameters, and communicates with the operator when displayed messages indicate problems.

Possible error categories associated with these human components and their tasks include:

1. *Operator errors:* Errors of memory, errors of identification and interpretation, and the cumulative impact of errors of operation [6,7].

2. *Link controller errors:* Errors of attention, errors of interpretation, errors of procedure and documentation, and errors of communication.

The levels of maintenance associated with the systems under study include the following:

1. *Level 1 maintenance:* This is performed mainly by the link controller following an initial diagnosis and report of anomalous condition by the operator. It requires the controller to provide a sequence of inputs or directives contained within an application software to accomplish system reconfiguration and linking of a standby unit into the operating path (i.e., system restoration). The reconfiguration and link completion time is an indicator of the controller's level of performance.

2. *Level 2 maintenance:* This is performed on site by maintenance personnel equipped with diagnostic and test capabilities. It involves diagnosis and fault isolation within the malfunctioning unit, removal and replacement of the faulty component, and verification of proper functioning of the repaired unit.

III. Data-Processing Systems Reliability Definitions and Measures

Most reliability definitions and concepts used in the classical reliability area are applicable to data-processing systems analysis. However, there are a few terms that are particularly pertinent. A brief description of these terms (to appear in the model) follows.

The *system outage* occurs when both units in the standby configuration are disconnected for repair from various failure modes.

A *temporary outage* occurs when one unit has failed and the link controller is involved in the reconfiguration and linking task to restore system operation.

Degraded refers to a unit's performance characterized by a full-capacity throughput associated with a higher error rate per data block.

System degraded occurs when an operating unit has become degraded for which the standby unit is inoperable.

Temporary degraded occurs when the operating unit is in a degraded mode and the link controller is involved in restoring system operation.

An integrated operational analysis of the data systems under study requires that all their aspects of behavior, which have been described earlier, need to be considered. Therefore, formal reliability analysis and modeling of these systems have been conducted and some of the relevant performance measures have been derived. Satisfactory system performance from the customer's point of view, namely, that of the using spacecraft, is affected by frequencies and durations of outages and degraded output. Thus, adequate measures of reliability may include the proportions of normal data and degraded data acquired, "combined" data availability, and the probability of system outage due to human errors. In the subsequent sections, the Markov process technique is employed to derive exact expressions for some of these performance measures.

IV. Model Description and Assumptions

The following are the specific assumptions of the proposed Markov model:

1. The model represents a two-unit identical standby redundant system with operator-initiated failure and degraded states.
2. The operating unit may fail due to either a hardware fault or an operator error. The unit may also operate in a degraded mode due to partial hardware failures and/or noncritical operator errors.
3. Various failure modes are statistically independent.
4. Hardware and operator-initiated failure rates, as well as the degradation rate, are all constants.
5. The link controller performs the reconfiguration and linking process. The completion time of this process is exponentially distributed.
6. The physical repair of the failed or degraded unit (level-2 maintenance) is undertaken as soon as the reconfiguration and linking task is completed. The repaired unit is recovered directly to the fully operable or operating mode.
7. Repair rates (i.e., level-2 maintenance completion rates) from various failure modes are constants.
8. System operating policies require that the degraded unit of the system must not be repaired at its degraded output during the linking process of the standby unit.
9. The degraded unit of the system is regarded as failed (due to hardware or human-initiated failures) when providing degraded-unusable output; that is, when the error rate per data block is greater than a previously specified criterion or tolerance limit.
10. The failure rate of a unit in its standby mode is zero.
11. A unit may exist in one of three possible states—operating, degraded, or failed (with total output loss).
12. Initially, the system is in the fully operating state with one unit being in an operable standby mode.
13. During system operation, only one change can take place in the state of the system at each instantaneous change or increment of time.

V. Model Development

A. Notation List

The state-space diagram associated with the two-unit standby human-machine system is shown in Fig. 1. The following symbols are associated with this system model:

S_i denotes the ith state of the system, for $i = 0, 1, 2, \cdots, 15$; $i = 0$ (one unit is operating, the other unit is in standby mode); $i = 1$ (one unit has failed due to a hardware failure); $i = 2$ (one unit failed due to an operator error); $i = 3$ (one unit is in a degraded state); $i = 4, 5, 6$ (one unit is linked, the other unit is under repair from a failure mode); $i = 7, 8, 9$ (one unit is under repair from a failure mode, the other unit is in a degraded state); $i = 10$ (both units are under repair due to hardware failures); $i = 11, 12$ (both units are under repair—one from a hardware

failure and the other from an operator-initiated failure); $i = 13$ (both units are under repair from operator-initiated failures); $i = 14, 15$ (both units are under repair—one from a degraded mode, the other from a hardware failure and operator-initiated failure, respectively)

$P_i(t)$ denotes the probability that the system is in state S_i at time t, for $i = 0, 1, 2, \cdots, 15$.

λ_j is the constant failure or degraded rate of a unit due to the jth failure mode, where $j = 1, 2, 3$, corresponding to hardware failure, operator-initiated failure, and degraded modes, respectively.

λ'_j is the constant failure rate of the degraded unit due to the jth failure mode, where $j = 1, 2$, corresponding to hardware and human-initiated failure modes, respectively.

λ_{Lj} is the link controller's completion rate of the linking task of the standby unit given the jth failure mode, where $j = 1, 2, 3$, corresponding to hardware failure, operator-initiated failure, and degraded failure modes, respectively. That is, the link controller's completion rate of the level-1 maintenance task.

μ_j is the constant repair rate of a failed unit from the jth failure mode, where $j = 1, 2, 3$, corresponding to the hardware failure, operator-initiated failure, and degraded modes, respectively.

μ_4, μ_5 are the constant repair rates of either one of the two failed units from hardware failure and operator-initiated failure modes, respectively, that is, $\mu_4 = 2\mu_1$ and $\mu_5 = 2\mu_2$.

B. Model Formulation

The following differential equations were developed by examining the discrete-space stochastic process described in Fig. 1:

$$P'_0(t) = -(\lambda_1+\lambda_2+\lambda_3)P_0(t)+\mu_1 P_4(t)+\mu_2 P_5(t)+\mu_3 P_6(t)$$

$$P'_1(t) = \lambda_1 P_0(t) - \lambda_{L1} P_1(t) + \lambda'_1 P_3(t)$$

$$P'_2(t) = \lambda_2 P_0(t) - \lambda_{L2} P_2(t) + \lambda'_2 P_3(t)$$

$$P'_3(t) = \lambda_3 P_0(t) - (\lambda_{L3} + \lambda'_1 + \lambda'_2) P_3(t) + \mu_1 P_7(t) + \mu_2 P_8(t)+\mu_3 P_9(t)$$

$$P'_4(t) = \lambda_{L1} P_1(t) - (\lambda_1 + \lambda_2 + \lambda_3 + \mu_1) P_4(t) + \mu_4 P_{10}(t) + \mu_2 P_{12}(t)+\mu_3 P_{14}(t)$$

$$P'_5(t) = \lambda_{L2} P_2(t) - (\lambda_1 + \lambda_2 + \lambda_3 + \mu_2) P_5(t) + \mu_1 P_{11}(t) + \mu_5 P_{13}(t)+\mu_3 P_{15}(t)$$

$$P'_6(t) = \lambda_{L3} P_3(t) - (\lambda_1 + \lambda_2 + \lambda_3 + \mu_3) P_6(t) + \mu_1 P_{14}(t) + \mu_2 P_{15}(t)$$

$$P'_7(t) = \lambda_3 P_4(t) - (\lambda'_1 + \lambda'_2 + \mu_1) P_7(t)$$

$$P'_8(t) = \lambda_3 P_5(t) - (\lambda'_1 + \lambda'_2 + \mu_2) P_8(t)$$

$$P'_9(t) = \lambda_3 P_6(t) - (\lambda'_1 + \lambda'_2 + \mu_3) P_9(t)$$

$$P'_{10}(t) = \lambda_1 P_4(t) + \lambda'_1 P_7(t) - \mu_4 P_{10}(t)$$

$$P'_{11}(t) = \lambda_1 P_5(t) + \lambda'_1 P_8(t) - \mu_1 P_{11}(t)$$

$$P'_{12}(t) = \lambda_2 P_4(t) + \lambda'_2 P_7(t) - \mu_2 P_{12}(t)$$

$$P'_{13}(t) = \lambda_2 P_5(t) + \lambda'_2 P_8(t) - \mu_5 P_{13}(t)$$

$$P'_{14}(t) = \lambda_1 P_6(t) + \lambda'_1 P_9(t) - (\mu_1 + \mu_3) P_{14}(t)$$

$$P'_{15}(t) = \lambda_2 P_6(t) + \lambda'_2 P_9(t) - (\mu_2 + \mu_3) P_{15}(t) \quad (1)$$

with the following initial conditions at time $t = 0$:

$$P_0(0) = 1, \qquad P_i(0) = 0, \qquad \text{for } i = 1, 2, \cdots, 15 \quad (2)$$

VI. Solution of the Model

Considering the limiting or steady-state probabilities, the derivatives in the first set of equations (1) are set to zero. The relationship, $P_0 + P_1 + \cdots + P_{15} = 1$ is then utilized to obtain the following steady-state expressions:

$$P_i = K_i P_0, \qquad \text{for } i = 1, 2, \cdots, 15 \quad (3)$$

and

$$P_0 = \frac{1}{1 + K_1 + K_2 + \cdots + K_{15}} \tag{4}$$

where K_i is a constant associated with state i and defined as follows:

$$K_1 = \frac{\lambda_1 + \lambda_1' K_3}{\lambda_{L1}}, \quad K_2 = \frac{\lambda_2 + \lambda_2' K_3}{\lambda_{L2}}$$

$$K_3 = \left(\frac{a b_4 b_6 + \lambda_1 \mu_1 b_6 + \lambda_2 \mu_2 b_4}{m_1 b_4 b_6 + \mu_1 m_3 b_6 + \mu_2 m_4 b_4} \right)$$

$$K_4 = \frac{-(\lambda_1 + m_3 K_3)}{b_4}, \quad K_5 = \frac{-(\lambda_2 + m_4 K_3)}{b_6}$$

$$K_6 = \frac{-\lambda_{L3} K_3}{b_8}, \quad K_7 = \frac{\lambda_3 K_4}{a_5}$$

$$K_8 = \frac{\lambda_3 K_5}{a_6}, \quad K_9 = \frac{\lambda_3 K_6}{a_7}$$

$$K_{10} = \left(\frac{\lambda_1}{\mu_4} + \frac{\lambda_3 \lambda_1'}{a_5 \mu_4} \right) K_4, \quad K_{11} = \left(\frac{\lambda_1}{\mu_1} + \frac{\lambda_3 \lambda_1'}{a_6 \mu_1} \right) K_5$$

$$K_{12} = \left(\frac{\lambda_2}{\mu_2} + \frac{\lambda_3 \lambda_2'}{a_5 \mu_2} \right) K_4, \quad K_{13} = \left(\frac{\lambda_2}{\mu_5} + \frac{\lambda_3 \lambda_2'}{a_6 \mu_5} \right) K_5$$

$$K_{14} = \left(\frac{\lambda_1}{a_8} + \frac{\lambda_3 \lambda_1'}{a_7 a_8} \right) K_6, \quad K_{15} = \left(\frac{\lambda_2}{a_9} + \frac{\lambda_3 \lambda_2'}{a_9 a_7} \right) K_6$$

$$m_1 = \frac{\lambda_{L3} \mu_3}{b_8}, \quad m_2 = \frac{a_1 b_8 + \lambda_{L3} b_3}{b_8}$$

$$m_3 = \frac{\lambda_1' b_8 - \lambda_{L3} b_5}{b_8}, \quad m_4 = \frac{\lambda_2' b_8 - \lambda_{L3} b_7}{b_8}$$

$$b_1 = \frac{\lambda_3 \mu_1}{a_5}, \quad b_2 = \frac{\lambda_3 \mu_2}{a_6}, \quad b_3 = \frac{\mu_3 \lambda_3}{a_7}$$

$$b_4 = \frac{\lambda_3 \lambda_1' + \lambda_3 \lambda_2' + \lambda_1 a_5 + \lambda_2 a_5 - a_2 a_5}{a_5}$$

$$b_5 = \frac{\lambda_1 \mu_3 a_7 + \lambda_3 \lambda_1' \mu_3}{a_7 a_8}$$

$$b_6 = \frac{\lambda_3 \lambda_1' + \lambda_3 \lambda_2' + \lambda_1 a_6 + \lambda_2 a_6 - a_3 a_6}{a_6}$$

$$b_7 = \frac{\lambda_2 \mu_3 a_7 + \mu_3 \lambda_3 \lambda_2'}{a_7 a_9}$$

$$b_8 = \frac{(\mu_2 \lambda_3 \lambda_2' + \mu_2 \lambda_2 a_7)}{a_7 a_9} + \frac{(\lambda_3 \lambda_1' \mu_1 + \mu_1 \lambda_1 a_7)}{a_7 a_8} - a_4$$

$$a = \lambda_1 + \lambda_2 + \lambda_3, \quad a_1 = \lambda_{L3} + \lambda_1' + \lambda_2', \quad a_2 = a + \mu_1$$

$$a_3 = a + \mu_2, \quad a_4 = a + \mu_3, \quad a_5 = \lambda_1' + \lambda_2' + \mu$$

$$a_6 = \lambda_1' + \lambda_2' + \mu_2, \quad a_7 = \lambda_1' + \lambda_2' + \mu_3, \quad a_8 = \mu_1 + \mu_3$$

$$a_9 = \mu_2 + \mu_3$$

The steady-state probabilities given in Eqs. (3) and (4) can be used to develop system availability and reliability measures. These measures include:

1. Integrated steady-state availability denoted by A_S. This measure combines both normal and degraded availabilities and is given by

$$A_S = (1 + K_3 + K_4 + K_5 + K_6 + K_7 + K_8 + K_9) P_0 \tag{5}$$

2. The proportion of normal-performance, P_N, is given by

$$P_N = (1 + K_4 + K_5 + K_6) P_0 \tag{6}$$

3. The proportion of degraded-performance, P_D, is given by

$$P_D = (K_3 + K_7 + K_8 + K_9) P_0 \tag{7}$$

4. The probability of system failure (data outage) due to the cumulative impact of human errors, P_H, is given by

$$P_H = (K_{12} + K_{13} + K_{15}) P_0 \tag{8}$$

Equations (5) through (8) define the proposed availability model of the human–machine standby processing system.

VII. Application and Analysis

The steady-state availability model described in Eqs. (5) through (8) was implemented to evaluate the "operational" performance of several standby data-processing systems. The receiver, subcarrier demodulator assembly, and telemetry processing assembly (TPA) are examples of these systems. They are utilized in spacecraft-to-station communications at each Deep Space Communications Complex. The following two applications demonstrate the usefulness of some of these measures for performance analysis of a TPA assembly as a selected configuration.

A. Steady-State Availability Analysis

For the application of the availability measure given by Eq. (5) to the TPA assembly, integrated data availability was computed for the following specified parametric values:

$\lambda_1 = 0.02$ failure/hr, $\mu_1 = 0.3$ repair/hr

$\lambda_2 = 0.002$ to 0.03 failure/hr, $\mu_2 = 0.3$ repair/hr

$\lambda_3 = 0.006$ failure/hr, $\mu_3 = 0.2$ repair/hr

$\lambda_1' = 0.002$ failure/hr, $\mu_4 = 2\mu_1$

$\lambda_2' = 0.003$ failure/hr, $\mu_5 = 2\mu_2$

$\lambda_{L1} = 3.0$ link/hr by link controller

$\lambda_{L2} = 0.5$ to 3.0 link/hr by link controller

$\lambda_{L3} = 4.0$ link/hr by link controller

The numerical results pertaining to Eq. (5) indicate variation in the TPA system availability, which is described in Fig. 2. The plots demonstrate that TPA-integrated data availability decreases with increasing levels of operator-initiated outage rate λ_2 and decreasing levels of the corresponding completion rate of reconfiguration and linking λ_{L2}.

If the TPA operational availability requirement is 98.5 percent, then possible combinations of hardware and human reliability requirements include:

$\lambda_2 = 0.002$, $\lambda_{L2} = 1$, $\lambda_1 = 0.02$, and other parameters remain the same

$\lambda_2 = 0.004$, $\lambda_{L2} = 2$, $\lambda_1 = 0.02$, and other parameters remain the same

Thus, operational availability can be achieved and/or improved through a proper trade-off among operator error rate, link completion rate (or level-1 maintenance completion rate), and hardware failure rate. The optimal combination of reliability parameters (to be considered for future design reviews) is usually selected based on whether improvements are needed in the hardware area, operating personnel area, or both.

B. Degraded-Performance Analysis

The proportion of the degraded-performance measure given by Eq. (7) is applied to the TPA system by using the previously specified parametric values, except for the following parameters:

$\lambda_2 = 0.03$ failure/hr, $\lambda_{L1} = \lambda_{L2} = 4.0$ link/hr

$\lambda_3 = 0.002$ to 0.03 failure/hr,

$\lambda_{L3} = 0.5$ to 3.0 link/hr

The pertinent numerical results indicate variation in the TPA annual degraded duration, which is plotted in Fig. 3. The plots demonstrate how system-degraded duration increases when a unit partial failure rate λ_3 increases, and the corresponding completion rate of reconfiguration and linking λ_{L3} decreases.

If the TPA annual degraded throughput is desired not to exceed a total of 200 hours, then possible combinations of the required partial failure rate (i.e., degradation rate) and controller's completion rate of the reconfiguration task include, where other parameters remain the same:

$\lambda_3 \leq 0.006$, $\lambda_{L3} \geq 0.5$

$\lambda_3 \leq 0.010$, $\lambda_{L3} \geq 1.0$

When system-degraded performance occurs, the reconfiguration and linking process is usually performed carefully or delayed for some time (provided that data throughput remains usable-degraded); therefore, λ_{L3} is most frequently less than or equal to 1 link/hr. Thus, greater attention should be given to maintain λ_3 below 0.010 failure/hr. This can be achieved through improvements of assembly design and its human interface and operational characteristics in order to reduce frequencies of partial failures and noncritical operating errors.

The other system performance measures, including the proportion of normal-performance and the probability of data outage due to human errors, are given by Eqs. (6) and (8), respectively. These measures can be applied to the TPA system in a similar fashion by using selected parametric values. The variation in each measure as compared

with the changing levels of λ_2 and λ_{L2} can also be graphically demonstrated. It should be obvious that the proportion of normal-performance data acquired P_N constitutes a significant portion of integrated (normal and degraded) data availability A_S. On the other hand, increasing operator error rate λ_2 results in a higher probability of system failure due to operator errors.

The graphical and parametric analyses of the proposed model, as described earlier, allow for presenting the practical aspects of the model and assist in the interpretation of the results.

VIII. Conclusions

This article presented a stochastic (Markovian) availability analysis of a class of human–machine communication standby systems, namely, data-processing systems. In the developed model, the possibility of a degraded-performance occurrence was considered and the human task of noninstantaneous restoration of system operation was incorporated. Therefore, the proposed model is more general and adequate for the systems under study than existing models.

The steady-state performance expressions derived in this study can be used in conjunction with appropriate plots to facilitate various failure mode sensitivities and trade-off analysis. Such analysis would support the decision-making process with regard to the optimal combination of the operator-error rate, the hardware failure rate, and the system restoration rate, which may result in a system "operational" availability that meets or exceeds a prespecified requirement.

Acknowledgments

The author would like to thank Joseph P. Goodwin of the TDA Mission Support and DSN Operations Office for his encouragement and support of this study. In addition, gratitude is due to N. Hoang for providing actual system performance data samples from the Discrepancy Report System to compute failure and restoration rates, and to the operating and maintenance personnel of the Goldstone Deep Space Communications Complex for the initial guidance and useful discussions related to the performance modeling of operational data systems.

References

[1] K. W. Lee, F. A. Tillman, and J. J. Higgins, "A Literature Survey of the Human Reliability Component in a Man–Machine System," *IEEE Trans. Reliability*, vol. R-37, pp. 24–34, 1988.

[2] G. W. Nieuwhof, "Short Note—Human Error," *Reliability Engineering*, vol. 6, pp. 191–192, 1983.

[3] B. S. Dhillon, *Human Reliability With Human Factors*, New York: Pergamon Press, 1986.

[4] D. Meister, "The Problem of Human-Initiated Failures," *Eighth National Symposium on Reliability and Quality Control*, pp. 469–476, 1962.

[5] L. H. Nawrocki, M. H. Strub, and R. M. Cecil, "Error Categorization and Analysis in Man–Computer Communication Systems," *IEEE Trans. Reliability*, vol. R-22, pp. 135–140, 1973.

[6] T. N. Issa, *Availability Prediction of Human-Equipment Data-Flow Systems With Application to NASA Deep Space Station*, Ph.D. dissertation, The Wichita State University, Wichita, Kansas, 1990.

[7] L. V. Rigby, "The Nature of Human Error," *Proc. Ann. Tech. Conf. of ASQC*, pp. 457–465, 1970.

[8] A. D. Swain, "Some Problems in the Measurement of Human Performance in Man–Machine Systems," *Human Factors*, vol. 6, pp. 687–700, 1964.

[9] J. A. Adams, "Issues in Human Reliability," *Human Factors*, vol. 24, pp. 1–10, 1982.

[10] J. A. Adams, *Human Factors Engineering*, New York: Macmillan, 1989.

[11] D. Meister, "Methods of Predicting Human Reliability in Man–Machine Systems," *Human Factors*, vol. 6, pp. 621–646, 1964.

[12] T. Regulinski and W. Askren, "Stochastic Modeling of Human Performance Effectiveness Functions," *Proc. Ann. Reliab. Maint. Symp.*, pp. 407–416, 1972.

[13] B. S. Dhillon and R. Misra, "Effect of Critical Human Error on System Reliability," *Reliability Engineering*, vol. 12, pp. 17–33, 1985.

[14] B. S. Dhillon and S. Rayapati, "Analysis of Redundant Systems With Human Errors," *Proc. Ann. Reliab. Maint. Symp.*, pp. 315–321, 1985.

[15] B. S. Dhillon and S. Rayapati, "Reliability Analysis of Non-Maintained Parallel Systems Subject to Hardware Failure and Human Error," *Microelectronics and Reliability*, vol. 25, pp. 111–122, 1985.

[16] B. S. Dhillon and S. Rayapati, "Human Error and Common-Cause Failure Modeling of Redundant Systems," *Microelectronics and Reliability*, vol. 26, pp. 1139–1162, 1986.

[17] B. S. Dhillon and S. Rayapati, "Human Error and Common-Cause Failure Modeling of Standby Systems," *Maint. Manag. Internat.*, vol. 7, pp. 93–110, 1988.

[18] P. P. Gupta and A. Kumar, "Reliability and MTTF Analysis of a Non-Repairable Parallel Redundant Complex System Under Hardware and Human Failures," *Microelectronics and Reliability*, vol. 26, pp. 229–234, 1986.

[19] P. P. Gupta and A. Kumar, "Stochastic Behavior of a Two-Unit Cold Standby Redundant Electronic Equipment Under Human Failure," *Microelectronics and Reliability*, vol. 27, pp. 15–18, 1987.

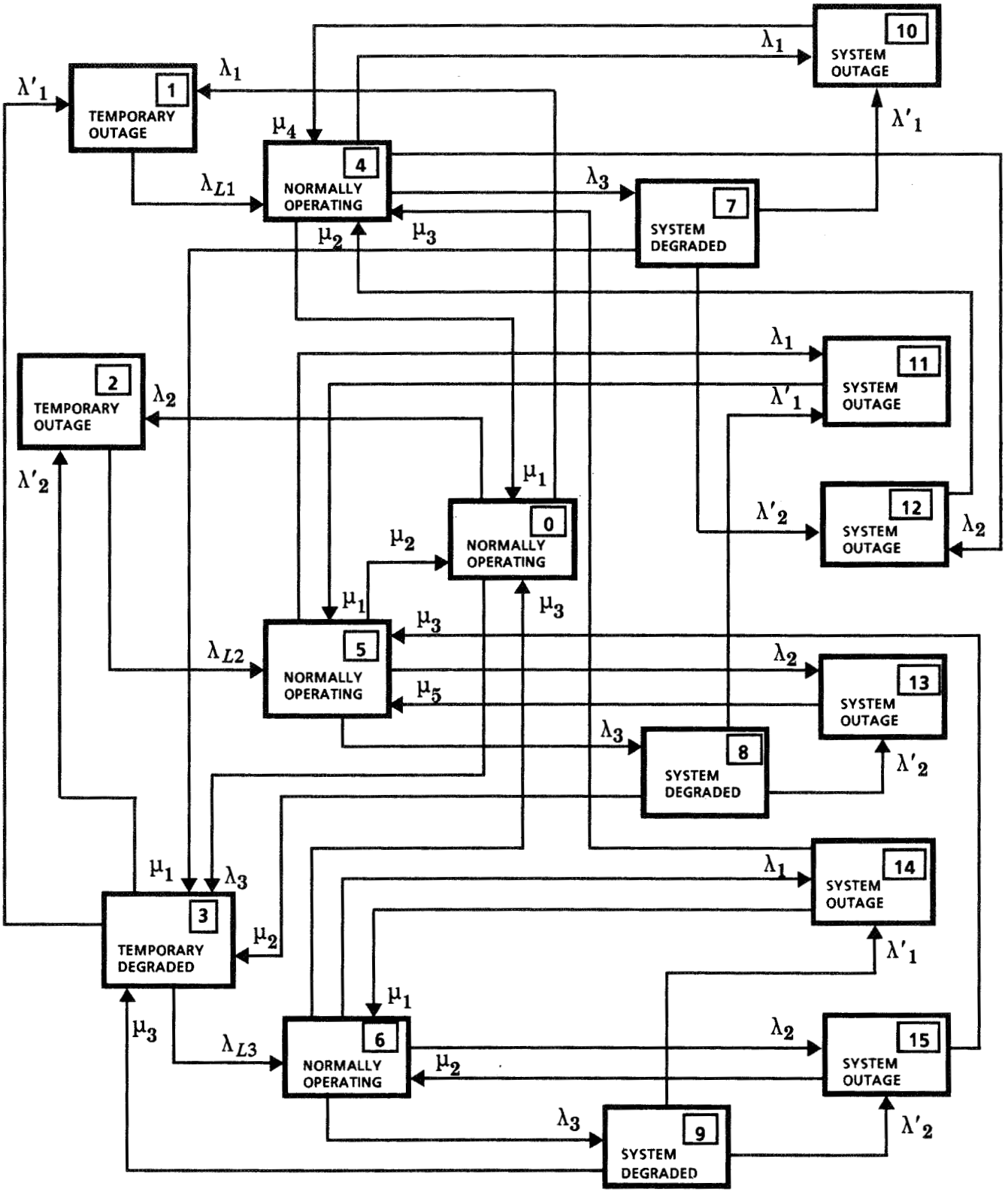


Fig. 1. State-space diagram of system stochastic performance.

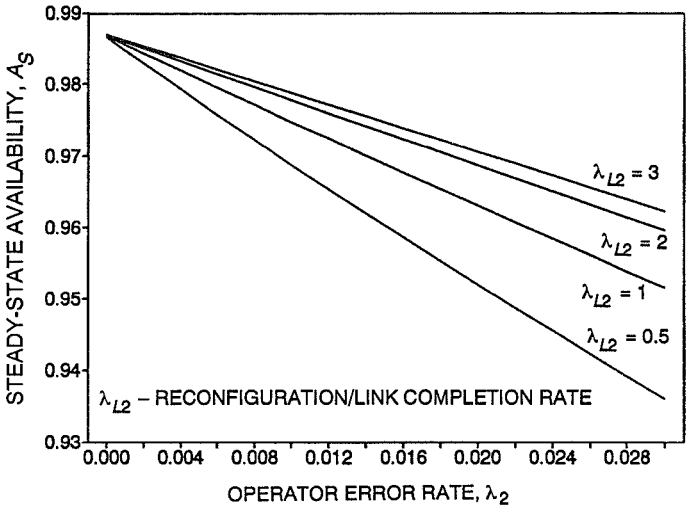


Fig. 2. TPA integrated data availability versus operator error rate.

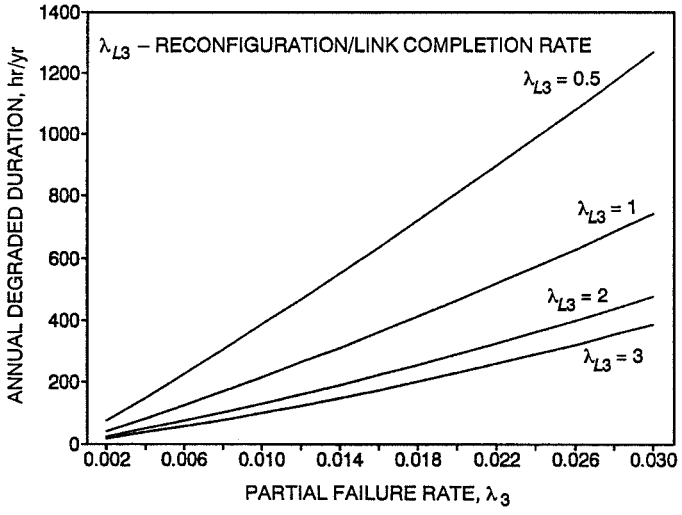


Fig. 3. TPA annual degraded performance versus partial failure rate.

Modeling Preparation Costs for Space Missions by Using Major Cost Drivers

J. S. Sherif
Software Product Assurance Section
and
California State University at Fullerton

D. S. Remer
TDA Planning
and
Harvey Mudd College of Engineering and Science

H. R. Buchanan
Radio Frequency and Microwave Subsystems Section

This article suggests a model for making long-range-planning cost estimates for DSN support of future space missions. The model is a function of major mission-cost drivers, such as maintenance and operations, downlink frequency upgrade, uplink frequency upgrade, telemetry upgrade, antenna gain/noise temperature, radiometric accuracy upgrade, radio science upgrade, and very long baseline interferometry.

The model is derived from actual cost data from three space missions: Voyager (Uranus), Voyager (Neptune), and Magellan. The model allows one to estimate the total cost and the cost over time of a similar future space mission.

The model was back-tested against the three projects—Voyager (U), Voyager (N), and Magellan—and gave cost estimates that range from 17 percent below to 19 percent above actual mission-preparation costs. The model was also compared with two other independent projects: Mariner Jupiter/Saturn (MJS later became Voyager) and Viking. The model gave total preparation-cost estimates that range from 15 percent above to 4 percent below actual total preparation costs for MJS and Viking, respectively.

I. Introduction

A. Project Objectives

The objective of this study is to develop a model that can be used in the early planning stages to estimate the cost to prepare the DSN for future space-mission support. The proposed model captures the major cost drivers of a mission, such as its use of an uplink, a downlink, a very long baseline interferometry system, etc. The proposed model gives cost estimates that are functions of the cost drivers and the duration of a project, which are demanded by the project's unique mission. The results of this study expand on previous cost modeling that was cost-driver and mission independent. The previous work can be used in the earliest stages of cost estimating, before the cost drivers and unique mission characteristics are defined [1].

The present study focuses on the major cost drivers that make up the total preparation cost of a particular project. Because of this focus, the total estimated project-preparation cost will reflect only those major cost drivers that pertain to that particular project, and thus a more project-sensitive cost estimate will be achieved.

B. Overview of Article

In Section II, the preparation cost drivers for space missions are defined, and the methodology for collecting the cost data and the cost history is summarized. The portion of the model concerning the total mission's cost-preparation drivers (Model A) is developed in this Section. Model A is back-tested against the three missions [Voyager (Uranus), Voyager (Neptune) and Magellan], and an example is given to show how to use it. In Section III, the aspect of the model concerning the cost drivers over a specific time period (Model B) is developed. Model B is back-tested against the three missions, and an example is given to show how to use it. In Section IV, as an "external" check, Model A is compared with two independent projects: Mariner Jupiter/Saturn (MJS, which later became Voyager) and Viking.

II. Development of a Model Based on Cost Drivers

A. Definition of "Cost Drivers"

The DSN-preparation costs for each project have been collected into the following major cost-driver categories:

(1) (M/O): maintenance and operations

(2) (D/L): downlink frequency upgrade

(3) (U/L): uplink frequency upgrade

(4) (TEL): telemetry upgrade

(5) (G/T): antenna gain/noise temperature

(6) (R/M): radiometric accuracy upgrade

(7) (R/S): radio science upgrade

(8) (VLBI): very long baseline interferometry

(9) (OTH): other

Here are the definitions of these cost drivers:

1. (M/O). These are initial entry and management costs for the project, e.g., funding for (M/O) network functions, network-operations project support, and the Tracking and Data System (TDS) manager for the project.

2. (D/L). These are the costs of adding new receiver, antenna, and microwave capabilities to the DSN: providing a new downlink frequency; giving additional performance capability to the existing receivers, antennas, and microwave instruments; increasing the number of channels provided by the existing antenna, receivers, and microwave instruments; etc.

3. (U/L). These are the costs of adding new transmitter, antenna, and microwave capabilities to the DSN: a new uplink frequency; additional performance capabilities to the existing transmitters, antennas, and microwave instruments, such as higher power, increased phase stability, etc.

4. (TEL). These are the costs of upgrading the telemetry and signal-processing equipment: adding new technical capability; adding to the monitor and control capability; providing new techniques, such as baseband combining for antenna arrays; etc.

5. (G/T). These are the costs of upgrading the ratio of the antenna gain to the receiving system noise temperature. (G/T) is a figure of merit for a telecommunications receiving system. Included are costs of providing new antennas, enlarging existing antennas, providing new/improved low-noise microwave amplifiers, providing antenna arrays, etc.

6. (R/M). These are the costs associated with upgrading the accuracy with which the spacecraft location can be measured. This includes upgrades to the data system equipment, improving DSN station location accuracy, improving time-synchronizing calibration throughout the network stations, etc.

7. (R/S). These are costs associated with upgrading the DSN radio science performance. These include

adding new and/or improved receivers and data-processing and recording equipment; improved frequency and timing equipment/calibration, etc.

8. (VLBI). These are costs of implementing new, complete (VLBI) equipment for both 34-m Wide Channel Bandwidth (WCB) and 70-m Narrow Channel Bandwidth (NCB) systems. Included are receivers, low-noise amplifiers, and support for the Radio Source Catalog and Universal Time Engineering.

9. (OTH). These costs are for any miscellaneous tasks not fitting into one of the above cost-driver categories. In some cases, the costs allocated to a cost-driver category are so small as to be presumed to be of a miscellaneous nature. These costs have been placed in the (OTH) category in this article. See Section II.C for further details.

B. Data Collection and Summary

The annual cost obligations used in this article are taken from Telecommunications and Data Acquisition (TDA) Work Authorization Documents (WAD Obligations Performance Reports), and do not include construction of facilities (CoF) costs, spacecraft costs, transportation costs, and/or other logistics costs.[1] All costs used in this article are adjusted for inflation to 1987 dollars by using the NASA Inflation Index. The preparation costs, grouped into cost drivers, for three projects—Voyager (U), Voyager (N), and Magellan—and the typical cost-driver values are shown in Table 1. The periods for tracking the preparation costs considered in this article are 1982 through 1986 for Voyager (U), 1985 through 1988 for Voyager (N), and 1985 through 1988 for Magellan [1].

C. Development of Model A: The Total-Mission Cost-Drivers Model

The assumption behind Model A is that the total preparation cost for a mission can be estimated by the summation of the typical cost-driver values given in Table 1 that are relevant to that particular mission. A typical cost-driver value in the model is an effective average value that is calculated after assigning any cost-driver values in a particular category that are less than 15 percent of the maximum value to the "Other" cost-driver category. It is assumed that a cost value that is that low reflects miscellaneous changes to the system rather than a significant cost-driver upgrade. For example, in Table 1, the 46 $K value for the (R/S) of Magellan is less than 15 percent of the 5,484 $K (R/S) value of Voyager (N); therefore, this cost driver for Magellan is considered "Other." Consequently, the typical (R/S) cost-driver value will be the average of those of Voyager (U) and Voyager (N), or 3,429 $K. The numbers in brackets in Table 1 were handled this way. Note that Magellan has 770 $K of miscellaneous costs in addition to the (R/M), 33 $K, and the (R/S), 46 $K, included in the "Other" cost-driver category.

D. Back-Testing the Total-Mission Cost-Drivers Model

The total-mission cost-drivers model (Table 1) was compared with the actuals for the three missions: Voyager (U), Voyager (N), and Magellan, as shown in Table 2.

A comparison of the actual preparation costs and those same costs, as predicted by Model A, for the three missions is shown in Table 3.

The average mission-preparation cost of 34.8 $M, as estimated by the model, is the same as the actual average cost. However, the difference in predicting individual mission-preparation costs ranges from 17 percent below to 18.9 percent above actual costs for Voyager (N) and Voyager (U), respectively. The preparation costs estimated from the model are about 1.9 percent below the actual preparation costs for Magellan.

E. How To Use Model A (the Total-Mission Cost-Drivers Model)

Model A is developed from the historical cost data of three space missions: Voyager (U), Voyager (N), and Magellan; the values of like preparation cost drivers are averaged. For example, to estimate the total preparation cost for a mission that has the cost drivers (R/M), (R/S), (VLBI), (U/L), (TEL), and (G/T), one determines the actual cost-driver value for each mission (in Table 1) and averages them to get cost-driver values for the model. A summary is given in Table 4.

Model A gives the total preparation cost of a mission, but it does not give a profile of costs over time. Model B does give the cost profile over time.

III. Model B: The Cost Drivers Modeled Over Time

A. Development of a Model of Cost Drivers Over Time

The average preparation cost over time for each major cost driver is calculated for each of the three missions:

[1] "Obligations Performance Reports, 1982–1988," TDA work authorization documents (WADs) (internal documents), Jet Propulsion Laboratory, Pasadena, California.

Voyager (U), Voyager (N), and Magellan. The individual cost-driver data are then regressed over time, and the best-fit equation is chosen. The best-fit equations are shown in Table 5, where Y_t is the cost in year t, $(t = 1, 2, ..., n)$; t is the number of the year in the DSN preparation-cost life of the mission; n is the total number of years of the DSN preparation; and the total cost of the DSN preparation is $Y(\text{total}) = \Sigma Y_t$.

B. Analysis of Model B: The Cost Drivers Modeled Over Time

The best-fit equations for the cost-driver data are linear for (U/L) and (VLBI); quadratic for (TEL), (G/T), and (R/M); and cubic for (M/O), (D/L), and (R/S). The equations have a goodness of fit (R^2) that ranges from 81 to 100 percent [2]. Figures 1 through 6 show the actual costs for each cost driver and those predicted by the equations. (U/L) and (VLBI) cost-driver figures are not shown since as of 1988, there were only two data points for each cost driver. This is the cutoff year of the data collected for the previous report on this research [1].

C. Back-Testing Model B: The Cost Drivers Modeled Over Time

Model B was checked against the three missions: Voyager (U), Voyager (N), and Magellan. Table 6 shows the actual average annual preparation costs of the three missions and those costs as predicted by the model. Table 7 shows that the 33.6 $M average preparation cost for the three missions, as predicted by the model, is 1.2 $M below the actual average preparation cost of 34.8 $M. The difference is about 3.4 percent. However, the difference in predicting individual mission preparation costs ranges from 22.5 percent below to 17.5 percent above actual preparation costs—for Voyager (N) and Voyager (U), respectively. The difference in predicting preparation costs for Magellan is 5 percent.

For planning purposes, a model that gives an accurate cumulative preparation cost for the life of the project is needed. For example, Fig. 7 shows the cumulative actual preparation cost over time and the cumulative preparation cost predicted by the cost-driver model for the Magellan mission, while Fig. 8 shows the cumulative actual preparation cost over time and the cumulative preparation cost predicted by the cost-driver model for cost driver (TEL) of Magellan.

In addition to comparing the actual preparation costs with those predicted by the model for a specific project or cost driver, the actual average preparation cost is also compared with the model's average for all three projects. For example, Fig. 9 shows the actual average preparation cost of a mission over time and the average preparation cost predicted by the model, while Fig. 10 shows the cumulative actual average preparation cost of a mission over time and the average cumulative preparation cost predicted by the model.

D. How To Use Model B: The Cost Drivers Modeled Over Time

Example 1. To estimate the annual preparation costs for a mission of five years that incurs all nine cost drivers, one looks up the model's results in Table 6 and sums the costs predicted by the model. A summary is given in Table 8.

Example 2. To estimate the total preparation costs over time for a mission of five years' duration that has the following six cost drivers: (R/M), (R/S), (VLBI), (U/L), (TEL), and (G/T), one looks up the model's results in Table 6 and sums the costs predicted by the model. The results are shown in Table 9.

This total value (50.5 $M) is close to the value (50.3 $M) predicted by Model A—the total-mission cost-drivers model, as described in Section II.E. The difference is about 0.2 $M, or 0.4 percent above that predicted by Model A.

Example 3. To estimate the preparation costs for any cost driver, such as (TEL), over a period of five years, one looks up the model's results in Table 6 and obtains the results shown in Table 10. This technique allows one to see the cost profile for this cost driver over time, and the same technique could be used for all the cost drivers for a proposed project.

IV. External Check and Comparison With Independent Missions

Model A was tested against two other independent missions: Mariner Jupiter/Saturn (MJS) and Viking.

The project MJS seems to have incurred all the major cost drivers that are covered in this article.[2] According to Table 1, the total cost for an average mission with a four- to five-year duration that incurs all the major cost drivers is 56.1 $M (1987 $M). Based on a previous study [1], it was concluded that the MJS mission (which continued for 10 years) might be viewed as having two distinct phases. It seems reasonable to consider the cost estimate for such

[2] J. R. Hall, private communication, Jet Propulsion Laboratory, Pasadena, California, April 1990.

a mission as twice that predicted for the "standard" five-year mission. That results in a predicted cost of 112.2 $M (1987 $M), as compared with the actual cost of 97.5 $M.[3] The difference is about 15 percent.

The project Viking did not use (VLBI); however, it used the other cost drivers.[4] The estimated total preparation cost for Viking predicted by Model A is then 56,079 $K − 4,436 $K = 51,643 $K (1987 $K), or about 51.6 $M, as compared with the actual cost of 49.7 $M. The difference is 3.8 percent.

V. Summary

A cost model has been presented in this article to give estimates for future DSN-preparation costs. The model has two components: A and B.

Model A is called the total-mission cost-drivers model, and Model B models cost drivers over time.

Model A estimates total DSN-preparation costs based on the average values of DSN cost drivers from three space missions: Voyager (U), Voyager (N), and Magellan. The model is concerned with those cost drivers that are relevant to a mission, and thus, the model is sensitive to mission objectives and uniqueness. Model A does a reasonable job of representing the actual preparation costs for Voyager (U), Voyager (N), and Magellan. Based on back-testing of the actual three projects against the model, the results are in the range of 17 percent below to 19 percent above actual costs. Model A was also compared with two other independent projects, MJS and Viking. The model gave total-cost estimates that range from 15 percent above to 4 percent below actual total costs for MJS and Viking, respectively.

Model B estimates the annual preparation cost of each cost driver relevant to a mission and also estimates total mission-preparation cost. The model is time and cost-driver sensitive and thus will capture future missions' cost dependence on both time and relevant cost drivers. Model B also does a reasonable job of representing the actual preparation costs over time for Voyager (U), Voyager (N), and Magellan. Based on back-testing of the actual three projects against the model, the results are in the range of 22.5 percent below to 18 percent above actual costs.

Both Model A and Model B are applicable to missions that do not exceed five years' duration and that have the cost drivers discussed in this study.

[3] J. W. Layland, private communication, Jet Propulsion Laboratory, Pasadena, California, March 1990.

[4] D. J. Mudgway, private communication, Jet Propulsion Laboratory, Pasadena, California, April 1990.

References

[1] J. S. Sherif, D. S. Remer, and H. R. Buchanan, "Long-Range Planning Cost Model for Support of Future Space Missions by the Deep Space Network," *TDA Progress Report 42-101*, vol. January–March 1990, Jet Propulsion Laboratory, Pasadena, California, pp. 179–190, May 15, 1990.

[2] J. T. McClure and P. G. Benson, *Statistics* (4th ed.), New York: Macmillan Co., 1988.

Table 1. Preparation cost drivers for three missions, and typical cost-driver values, 1987 $K[a]

Cost Drivers	Missions			Typical Cost-Driver Value
	Voyager (U)	Voyager (N)	Magellan	
M/O	955	677	634	755
D/L	2,647	6,566	2,211	3,808
U/L	[700]	0	8,947	8,947
TEL	10,357	[1,134]	15,445	12,901
G/T	17,795	21,008	0	19,402
R/M	2,032	580	[33]	1,306
R/S	1,374	5,484	[46]	3,429
VLBI	0	[601]	4,436	4,436
Other	700	1,735	849	1,095
Total	35,860	36,050	32,522	56,079

[a]The bracketed numbers are discussed in Section II.C.

Table 2. Actual and Model A costs for three missions

Cost Drivers	Voyager (U)		Voyager (N)		Magellan	
	Actual	Model	Actual	Model	Actual	Model
M/O	955	755	677	755	634	755
R/M	2,032	1,306	580	1,306	0	0
R/S	1,374	3,429	5,484	3,429	0	0
D/L	2,647	3,808	6,566	3,808	2,211	3,808
VLBI	0	0	0	0	4,436	4,436
U/L	0	0	0	0	8,947	8,947
TEL	10,357	12,901	0	0	15,445	12,901
G/T	17,795	19,402	21,008	19,402	0	0
Other	700	1,095	1,735	1,095	849	1,095
Total	35,860	42,696	36,050	29,795	32,522	31,942

Table 3. Summary of actual and Model A total preparation costs in 1987 $M

Space Mission	Actual Preparation Cost, $M	Model A Preparation Cost, $M	Model A Minus Actual, Δ in $M	Error, %, Δ/Actual
Voyager (U)	35.9	42.7	6.8	18.9
Voyager (N)	36.0	29.8	−6.2	−17.0
Magellan	32.5	31.9	−0.6	−1.9
Average for all missions	34.8	34.8	0	0

Table 4. Preparation costs predicted by Model A in 1987 $M

Cost Driver	Cost Predicted by Model A
R/M	1.3
R/S	3.4
VLBI	4.4
U/L	8.9
TEL	12.9
G/T	19.4
Total	50.3

Table 5. The best-fit equation for each cost driver

Cost Drivers	Model	R^2
(M/O)	$Y_t = 352 - 318t + 144t^2 - 18.6t^3$	99
(D/L)	$Y_t = -3,053 + 4,558t - 1,474t^2 + 139t^3$	82
(U/L)	$Y_t = 4,561 - 25t$ (for $t = 3$ and 4)	100
(TEL)	$Y_t = -4,114 + 5,938t - 1,011t^2$	81
(G/T)	$Y_t = -2,175 + 5,880t - 1,053t^2$	99
(R/M)	$Y_t = 104 + 229t - 48.1t^2$	90
(R/S)	$Y_t = 1,339 - 2,207t + 1,218t^2 - 165t^3$	98
(VLBI)	$Y_t = -792 + 860t$ (for $t = 3$ and 4)	100

Table 6. Summary of actual average annual preparation costs for three missions and costs predicted by Model B, $K (1987 $). "A" = actual costs; "M" = Model B costs.

Year	M/O	D/L	U/L	TEL	G/T	R/M	R/S	VLBI	Other	Total
1 (A)	159	126	-	502	2,753	258	164	-	444	4,406
(M)	159	170	-	813	2,652	285	185	-	283	4,547
2 (A)	149	1,535	-	4,782	5,172	421	557	-	115	12,731
(M)	143	1,279	-	3,718	5,372	369	477	-	335	11,693
3 (A)	191	867	4,486	3,279	5,980	361	1,100	1,788	536	18,588
(M)	192	1,108	4,486	4,601	5,986	358	1,225	1,788	477	20,221
4 (A)	206	853	4,461	4,161	4,704	193	1,513	2,648	-	18,739
(M)	194	491	4,461	3,462	4,497	250	1,439	2,648	-	17,442
5 (A)	50	427	-	177	793	73	95	-	-	1,615
(M)	37	262	-	300	896	46	129	-	-	1,670
Total (A)	755	3,808	8,947	12,901	19,402	1,306	3,429	4,436	1,095	56,079
Total (M)	725	3,310	8,947	12,894	19,403	1,308	3,455	4,436	1,095	55,573
(M−A)	−30	−498	0	−7	1	2	26	0	0	−506

Table 7. Summary of the actual average annual preparation costs for three missions, and the costs predicted by Model B, $M (1987 $)

Space Mission	Actual Preparation Cost, $M	Model B Preparation Cost, $M	Model B Minus Actual, Δ in $M	Error, %, Δ/Actual
Voyager (U)	35.9	42.2	6.3	17.5
Voyager (N)	36.0	27.9	−8.1	−22.5
Magellan	32.5	30.8	−1.7	−5.0
Average for all missions	34.8	33.6	−1.2	−3.4

Table 8. Estimate of annual preparation costs for a five-year, nine-cost-driver mission by Model B, $K (1987 $)

Year	Actual Cost-Driver Total Cost, $K	Model B Cost-Driver Total Cost, $K	Model Minus Actual, Δ in $K	Error, %, Δ/Actual
1	4,406	4,547	141	3
2	12,731	11,693	−1,038	−8
3	18,588	20,221	1,633	9
4	18,739	17,442	−1,297	−7
5	1,615	1,670	55	3
Total	56,079	55,573	−506	0

Table 9. Estimate of annual preparation costs for a five-year, six-cost-driver mission by Model B, $M (1987 $)

Cost Driver	Cost Predicted by Model B
R/M	1.3
R/S	3.5
VLBI	4.4
U/L	8.9
TEL	13.0
G/T	19.4
Total	50.5

Table 10. Model B estimate of preparation cost of (TEL), $K (1987 $)

Year	Predicted (TEL) Cost, $K
1	813
2	3,718
3	4,601
4	3,462
5	300
Total	12,894

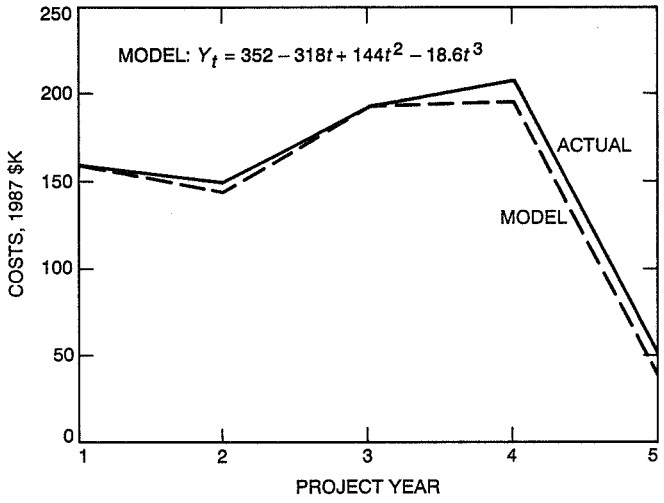


Fig. 1. Actual average maintenance and operations (M/O) costs versus costs predicted by Model B.

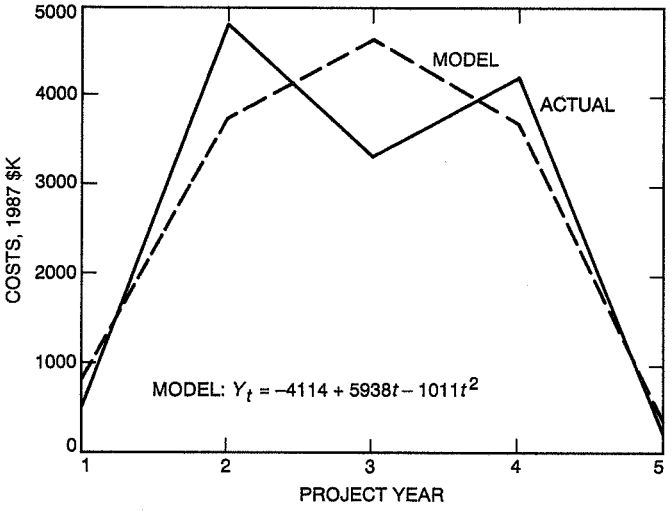


Fig. 3. Actual average telemetry upgrade (TEL) costs versus costs predicted by Model B.

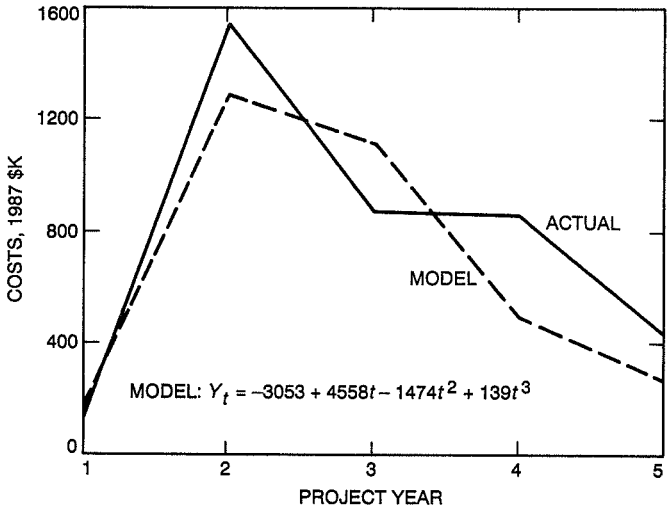


Fig. 2. Actual average downlink frequency upgrade (D/L) costs versus costs predicted by Model B.

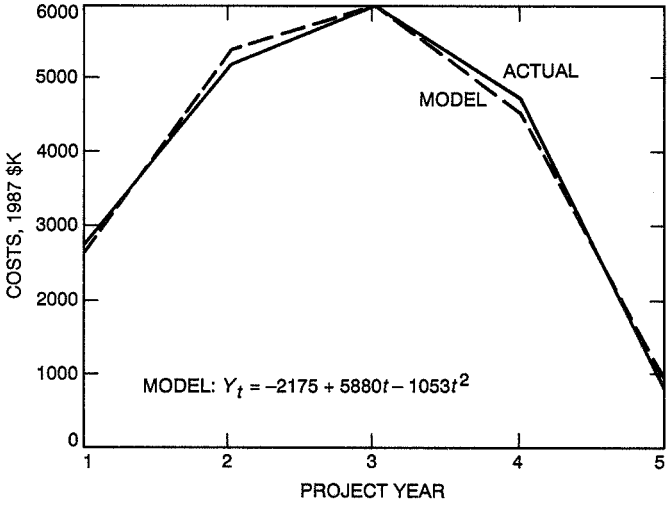


Fig. 4. Actual average antenna gain over system noise temperature (G/T) costs versus costs predicted by Model B.

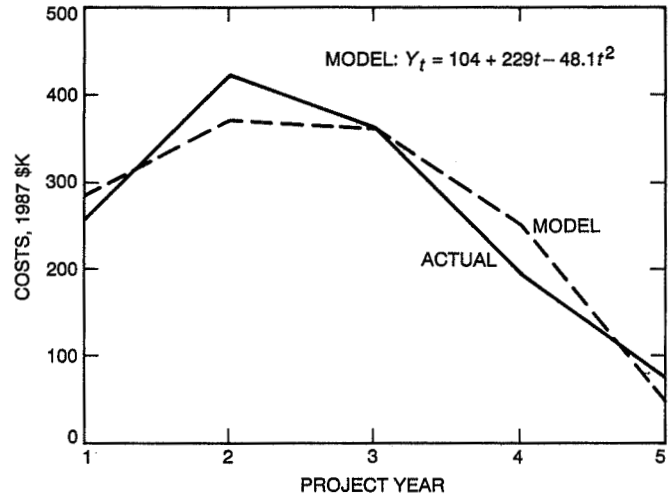


Fig. 5. Actual average radiometric accuracy upgrade (R/M) costs versus costs predicted by Model B.

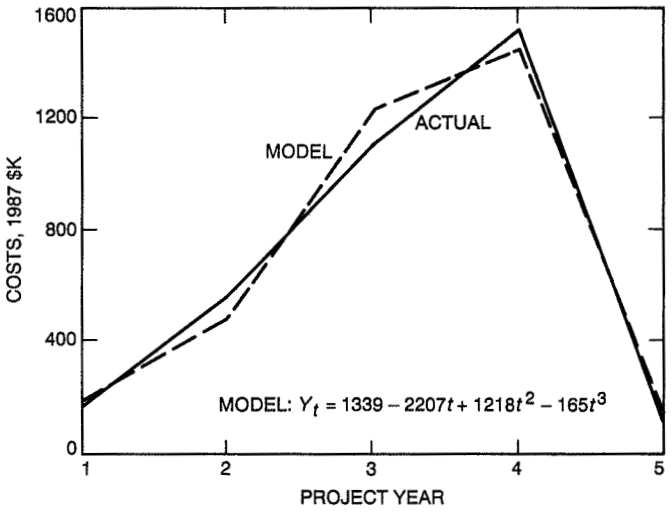


Fig. 6. Actual average radio science stability upgrade (R/S) costs versus costs predicted by Model B.

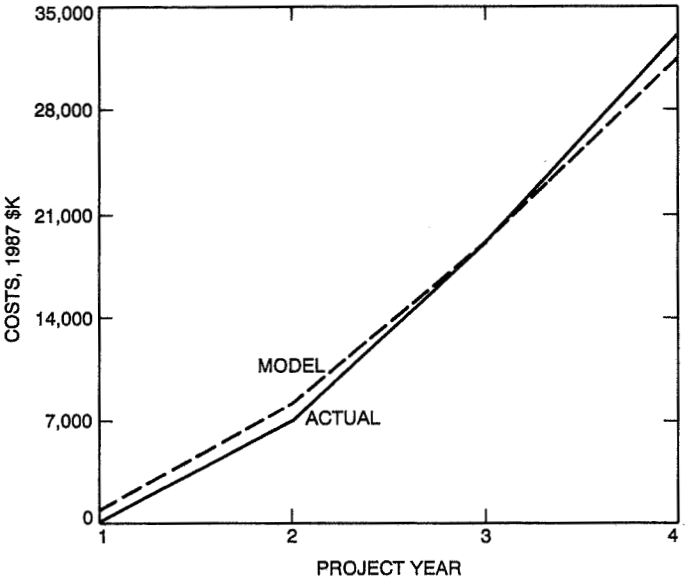


Fig. 7. Actual cumulative preparation costs and costs predicted by Model B for Magellan mission.

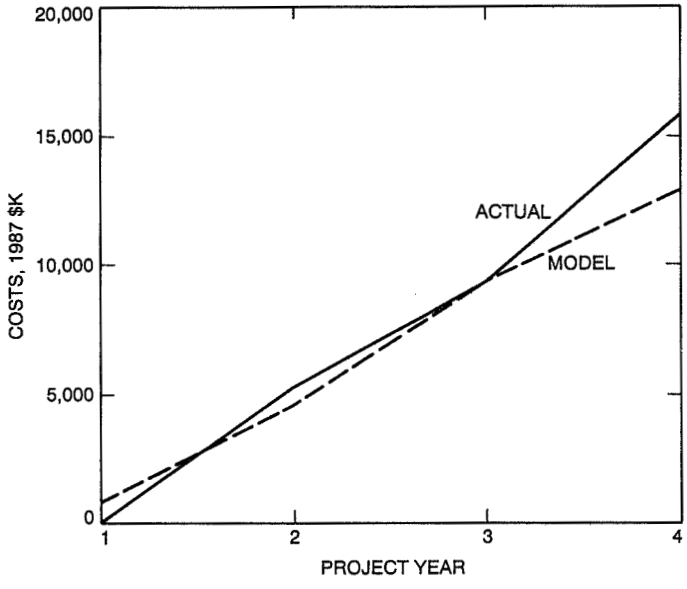


Fig. 8. Actual cumulative preparation costs and costs predicted by Model B for (TEL) cost driver of Magellan mission.

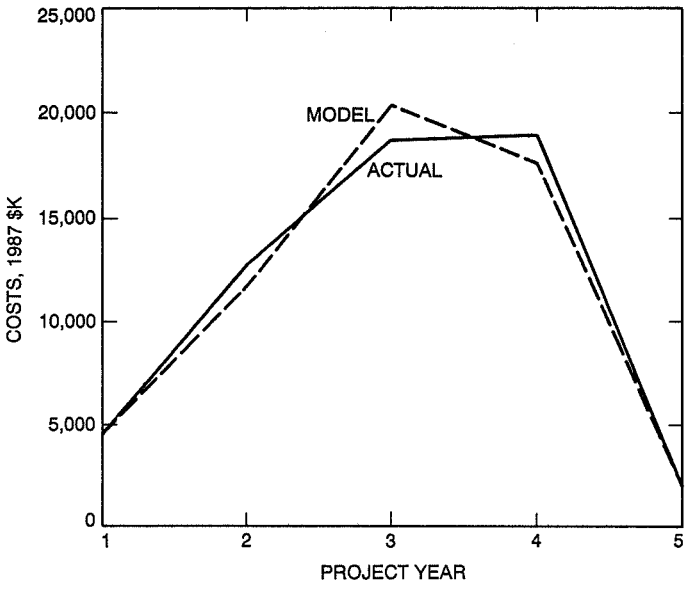


Fig. 9. Actual average preparation costs of a mission versus costs predicted by Model B.

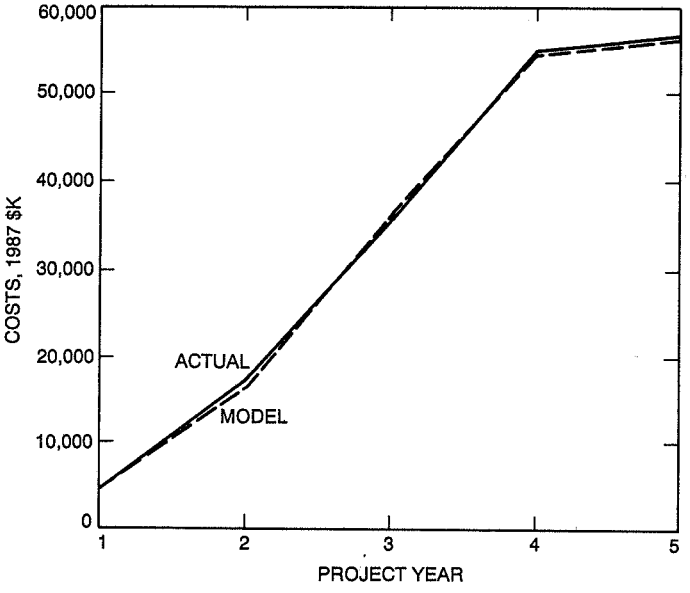


Fig. 10. Cumulative actual average annual preparation costs of a mission versus costs predicted by Model B.

Referees

The following people have refereed articles for *The Telecommunications and Data Acquisition Progress Report.* By attesting to the technical and archival value of the articles, they have helped to maintain the excellence of this publication during the past year.

H. Ansari
D. A. Bathker
J. J. Bautista
J. A. Berthias
R. S. Bhat
S. Butman
A. Cha
J. Chen
K.-M. Cheung
W. Chew
J. J. Cucchissi
R. J. Dewey
D. Divsalar
S. Dolinar
J. Ellis
D. M. Erickson
A. P. Freedman
M. Gatti
F. Y. Hadaegh
C. L. Hamilton
H. Hemmati
D. R. Hersey
J. K. Holmes
R. L. Horttor
W. J. Hurd
C. S. Jacobs
C. E. Johns
R. Jurgens
A. V. Kantak
P. W. Kinman
J. R. Lesh
R. Levy
F. Manshadi
M. McKenzie
X X Newhall
F. T. Nicholson
S. M. Petty
F. Pollara
E. C. Posner
J. Prestage
M. D. Rayman
D. L. Robinson
R. Sadr
B. Shah
M. Shahshahani
S. D. Slobin
E. M. Standish, Jr.
P. Stanton
J. I. Statman
R. Stevens
L.-Y. Sung
L. Swanson
M. Thorburn
S. W. Thurman
J. S. Ulvestad
K. Wilson
L. J. Wood
S.-C. Wu
G. Zimmerman